

Tartu Ülikool
Loodus- ja täppisteaduste valdkond
Ökoloogia ja maateaduste instituut
Geograafia osakond

Magistritöö geoinformaatikas

**Mobiilpositsioneerimise andmete ajalise jaotuse mõju mobiilsuse
parameetritele**

Taavi Pipar

Juhendaja: PhD Anto Aasa

Kaitsmisele lubatud:

Juhendaja:

Osakonna juhataja:

Tartu 2018

Annotatsioon

Mobiilpositsioneerimise andmete ajalise jaotuse mõju mobiilsuse parameetritele

Antud Magistritöö eesmärgiks on koostada metoodika YouSensi GNSS andmete ja mobiilpositsioneerimise andmete võrdlemiseks mobiilsusparameetrite abil. Töö hüpoteesiks on, et mobiilpositsioneerimise andmetest saadud indiviidi mobiilsusparameetrid on seda vähem kallutatud, mida ühtlasemalt on mobiilpositsioneerimise andmed jaotunud. Töös kasutati YouSense mobiilpositsioneerimis ja GNSS andmeid. Selgus, et ajaline entroopia on oluliselt setud ruumientroopia ja liikumisdistsantsi veaga kuid ajaline entroopia suurendab vea tekkimist. Ajalise entroopia mõju hindamine vajab kontrollimist ka suurema valimi ning passiivsete mobiilpositsioneerimise andmete põhjal.

Märksõnad: mobiilpositsioneerimine, GNSS, mobiilsusparameetrid

CERCS kood: P160 Statistika, operatsioonanalüüs, programmeerimine, finants- ja kindlustusmatemaatika

S230 Sotsiaalne geograafia

Impact of the time distribution of mobile positioning data on mobility parameters

The aim of this master's thesis was to develop a methodology for comparing the data of the YouSens GNSS and mobile positioning data using the mobility parameters. The hypothesis of this research is that the individual mobility parameters of mobile positioning data are less biased, the more consistently the mobile positioning data is distributed. YourSense mobile positioning and GNSS data were used in the research. The study revealed that temporal entropy is correlated positively with movement entropy and gyration of radius error and temporal entropy increases the error. The assessment of the time entropy impact needs to be verified on the basis of a larger sample and passive mobile positioning data.

Keywords: mobile positioning, GNSS, mobility parameters

CERCS code: P160 Statistics, operation research, programming, actuarial mathematics

S230 Social geography

Sisukord

Sissejuhatus	5
1. Teoreetilised lähtekohad	7
1.1 Passiivsete mobiilpositsioneerimise andmete valideerimine	7
1.1.1 Individuaalsete mobiilsusparameetrite valideerimine	7
1.2 Kõnetoimingutest tekkiv kallutatus	10
2. Andmed ja metoodika	12
2.1 Uurimisala	12
2.2 Andmestik	12
2.2.1 YouSense andmestik	12
2.2.2 OpenCellID mobiilivõrgu andmestik	14
3. Teekondade ja peatuspaikade eraldamine YouSense andmetest	15
3.1.1 Andmete puhastamine	15
3.1.2 Peatuspaikade tuvastamine YouSense GNSS andmetelt	15
3.1.3 Peatuspaikade grupeerimine	16
3.2 YouSense mobiilpositsioneerimise andmete ettevalmistus	18
3.3 Mobiilsusparameetrite arvutamine	18
3.3.1 Liikumisdistsants	18
3.3.2 Tegevusraadius	18
3.3.3 Ruumientroopia	19
3.3.4 Ajaline entroopia	19
3.4 Korrelatsioon ja regressioonanalüüs	20
4. Tulemused ja arutelu	21
4.1 Andmevead algandmetes	21
4.2 Liikumisdistsants	26
4.3 Ruumientroopia	29
4.4 Tegevusraadius	31

4.5	Lisaparametrid	32
4.6	Ruumientroopia prognoosjäägi standardhälbe regressioonanalüüs.....	34
5.	Kokkuvõte	40
6.	Impact of the time distribution of mobile positioning data on mobility parameters.....	41
7.	Tänuavaldused.....	42
	Kasutatud kirjandus.....	43
	Internetiallikad	45
	Lisa 1	46
	Lisa 2	47

Sissejuhatus

Traditsioonilisi rahvastiku-uuringuid asendatakse üha enam suurandmete (big data) põhjal tehtud analüüsidega. Suurandmete analüüsi eelistamise üheks põhjuseks on surve hoida uuringute hinda madalal ning võimalus koguda täpsemalt ja tihedamini statistilisi näitajaid. Suurandmed on tavaliselt tekkinud mingi muu digitaalse toimingu kõrvalt ning valim on juba moodustatud kasutajate enda aktiivsusest. Tavaliselt puudub uurimuse koostajatel võimalus algse valimi suunamiseks. Sellest tulenevalt on uurimuse koostajatel kasutada juba olemasolev valim, mis võib mingites aspektides olla kallutatud, sest uurijatel puudub kontroll valimi koostamise üle. Üheks selliseks andmestikuks, mida kasutatakse, on mobiilpositsioneerimise andmed. Vaatamata sellele, et mobiilpositsioneerimise andmestik on suurandmestik, ei taga see veel andmete esinduslikkust. Valimite moodustamisest on teada, et suurem valim tõstab tulemuste täpsust, kuid olenevalt valimi moodustamisest ei tõuse alati saadud andmete esinduslikkus. Ilma teadmisetä, kuidas ja millistel põhjustel valimivead on tekkinud, võib teha suurandmetest valesid järeldusi.

Varasemalt on uuritud mobiilpositsioneerimise andmete esinduslikkust (Chen et al., 2017a; Lu et al., 2017; Ranjan et al., 2012; Zhao et al., 2016), kus on vaadeldud, kuidas esinduslikkus sõltub mobiilikõnede keskmisest tihedusest. Varasemates uurimustes on kirjeldatud kõnetoimingute andmete valangulist (*bursty*) olemust ja sellest tulenevat võimalikku kallutatust (Blondel et al., 2015; Karsai et al., 2012; Ranjan et al., 2012). Ranjan et al. (2012) viitas, et uuringutes kasutatakse sageli aktiivsemaid mobiiltelefoni kasutajaid, kus vahel on piir isegi 0,5 kõnetoimingut tunnis, et tagada andmete täielikkus. Seega pole suurel hulgal rahvastikust võimalik osaleda mõningates uuringutes. Selline valimi koostamine meetod viib kallutatud tulemusteni. Ranjan et al. (2012) kasutas valangulisuse vähendamiseks 15 minutilisi vahemikke, eesmärgiga ühtlustada järjestikuseid kõnetoimingud, et vähendada kallutatust. Iovan et al. (2013) on kasutanud entroopiat kvaliteedi hindamiseks. Entroopia ehk korrapäratuse määr näitab juhuslike andmete keskmist informatsiooni hulka. Hetkel puudub autorile teadaolevalt ajalise entroopia näitaja kasutamine mobiilpositsioneerimise andmete töötlemisel. Ajalist entroopiat on võimalik kasutada kasutaja andmete kvaliteedi ja nende esinduslikkuse hindamiseks, mis võimaldab koostada uuringute jaoks vähem kallutatuid valimeid. Juhul kui kõnede tihedus ei mõjuta osade mobiilsusparameetrite täpsust, siis sarnane omadus võib esineda ka ajalisel entroopial.

Mitmetes esinduslikkuse uuringutes (Naboulsi et al., 2016; Zhao et al., 2016) on rõhutatud, et nende leitud seaduspärasused ei pruugi kehtida mõnes teises linnas või kultuuris. See on üks põhjustest, miks on vaja teostada sarnaseid uuringuid igas kultuuriruumis. Lisaks sellele, sõltub erinevate mobiilsusparameetrite esinduslikkus nende arvutamise metoodikast ning seda tuleks alati kontrollida. Kuna kasutatav andmestik on sõltuv kasutajate harjumustest, mis on ajas muutuv, tuleks teotada vastavaid uuringuid järjepidevalt.

Antud töö eesmärgiks on koostada metoodika YouSensi GNSS (globaalne satelliitnavigatsioonisüsteemi) andmete ja mobiilpositsioneerimise andmete võrdlemiseks mobiilsusparameetrite abil. Eesmärgi täitmiseks tuleb uurida mobiilpositsioneerimise andmete ajalise jaotuse mõju ruumi parameetritele ajalise entroopia abil. Töö hüpoteesiks on, et mobiilpositsioneerimise andmetest saadud indiviidi mobiilsusparameetrid on seda vähem kallutatud, mida ühtlasemalt on mobiilpositsioneerimise andmed jaotunud.

1. Teoreetilised lähtekohad

1.1 Passiivsete mobiilpositsioneerimise andmete valideerimine

Varasemalt on passiivsete mobiilpositsioneerimise andmete valideerimisel vaadeldud kolme suuremat väljundit: rahvastiku ruumiline jaotumine, agregeeritud liikumise vood ja individuaalsed mobiilsusparameetrid (Naboulsi et al., 2016). Lisaks sellele võivad väljundiks olla mobiilsuse mudelid, võrgustiku analüüs jne. Omavahel on võrrelnud rahvastiku ruumilist jaotust passiivse mobiilpositsioneerimise andmetest ning võrreldud seda rahvaloenduse andmetega (Calabrese et al., 2011; Frias-Martinez and Virseda, 2012; Pestre et al., 2016).

1.1.1 Individuaalsete mobiilsusparameetrite valideerimine

Naboulsi et al. (2016) on oma töös jaotanud passiivsete mobiilpositsioneerimise andmetest saadud individuaalsete mobiilsusparameetrite valideerimise kaheks: passiivsete mobiilpositsioneerimise andmete sisene valideerimine (*intra-CDR*) ning vaatlusandmestiku (*CDR-to-ground-truth*) abil valideeritavaks. Mõlema meetodi puhul on valideeritava andmestiku võrdluseks olev andmestik oma ruumi ja/või ajalise lahutusvõime poolest täpsem ning detailsem.

1.1.1.1 Individuaalsete mobiilsusparameetrite valideerimine sisemiselt

Passiivsete mobiilpositsioneerimise andmete saamiseks ei koguta mitte ainult väljuvaid kõnetoiminguid, vaid sellele lisaks sissetulevaid kõnetoiminguid, regulaarseid asukoha uuendusi, perioodilisi asukoha uuendusi, telefoni sisse- ja väljalülitamisi ning kõnetoimingu ajal toimuvaid mobiilimasti vahetusi. Enamik eelloetletud parameetritest ei vaja kasutaja poolset aktiivset telefoni kasutamist. Neid andmeid saab koguda ka siis, kui telefoni ei kasutata (Ranjan et al., 2012). Eelnevale lisaks on võimalik võrdluseks kasutada ka mobiilset andmeside andmestikku. Selle juures salvestatakse aeg ning iga alustatud andmeside vahetuse asukoht mobiilimasti täpsusega, sarnaselt kõnetoiminute andmetele.

Regulaarsed asukoha uuendused on põhjustatud mobiili liikumisest ning sellest tulenevalt vajadusest leida ümbruskonnast kõige parema signaaliga mobiilimaste. Sellisel juhul ei ole mobiil aktiivses kasutuses omaniku poolt. Perioodiline asukoha uuendamine toimub mobiilimasti ja telefoni vahel, kui kindla aja jooksul pole mobiil otsinud uut ühendust. Mobiilimasti kasutatakse saamaks sidet, kas mobiil asub jätkuvalt selle masti piirkonnas, et vajadusel toimuksid ühendused kiiresti. Näiteks: kui peaks tulema kõne või SMS, siis ei hakata otsima mobiili kõikjalt - ehk kõik mobiilimastid Eestis ei hakka tegema päringut, kas mobiil on nüüd just nende piirkonnas. Päring toimub pidevalt viimati oldud mobiilimastist. Eesmärgiks on suurendada toimingute kiirust ning vähendada mobiilimastide koormust.

Üheks esimeseks, kes vaatles passiivselt kogutud mobiilpositsioneerimise andmeid koos detailsema andmestikuga, oli González et al. (2008). Võrdluseks kasutas ta 206 kasutaja mobiili asukohti, mis olid kogutud aktiivse positsioneerimise teel iga kahe tunni tagant terve nädala jooksul. Kahe andmestiku võrdluses selgus, et asukoha muutumise distantssi jaotused jälgivad sama seaduspärasust. Seda järeldust kinnitab ka (Trestian et al., 2009). Kuigi González et al. (2008) jõuavad töös järeldusele, et kõnede vahelise aja varieeruvus ei mõjuta üldist asukoha muutumise jaotust, on töös siiski leitud, et asukoha muutus kõnede vahelisel ajal alla 4 tunni on piiratud maksimaalse kaugusega, mida kasutajatel on võimalik läbida antud ajavahemiku jooksul. Alates 8 tunnist kõnede vahel on asukoha muutuse piiravaks teguriks kasutaja tegevusraadius (*radius of gyration*). Lisaks võrreldi tegevusraadiuste jaotumist passiivsete mobiilpositsioneerimise andmete, mis koguti 6 kuu jooksul ning aktiivselt kogutud andmete vahel. Tulemused olid sarnased ning jälgisid lihtsustatud astmelise seaduspärasuse jaotumust (*truncated power-law distribution*).

Põhjalikumalt on passiivse mobiilpositsioneerimise andmeid valideeritud läbi anonüümsete kõnede, SMS-de ja andmeside andmete, mis on saadud mobiilside operaatori kõne andmebaasidest (*Per-Call Measurement Data (PCMD)*) (Ranjan et al., 2012). Andmeid koguti San Franciscos üle miljoni kasutajalt ühe kuu jooksul 2011. aasta juulis, kuid töös on kasutatud vaid kasutajad, kes lisaks kõnede ja SMS-idele kasutasid ka andmesidet. Kokku oli neid umbes pool miljonit. Kõige väiksema andmestiku osa moodustasid kõned ning suurima andmeside. Mobiilse andmeside hulga ning kõrge ajalise jaotumise üheks põhjuseks võib olla automatiseeritud rakendused, mis pidevalt ilma kasutajalt aktiivsust nõudmata, ühenduvad andmesidevõrku. Näiteks, kontrollitakse saabunud e-kirju. Töö eesmärgiks oli valideerida, kas aktiivsemate kõnetoimingute tegijate kasutamine detailsetes uuringutes tekitab kallutatust. Selleks võrreldi erinevatest andmestikest kõnesid, SMS ja mobiilsest

andmesidest saadud olulisi kohtasid, eriti kodu ja töökohta ning erinevaid mobiilsusparameetreid: ruumientroopia (*Shannon entropy*) ja tegevusraadius. Eelnevalt toodud uuringu kohaselt erinesid olulised paigad (*significant locations*) erinevate andmestike (SMS, kõned ja mobiilside) tulemustes oluliselt. Samas kodude ning töökohtade asukohad erinevate andmestike vahel ei erinenud üle 95% kasutajatest. Lisaks sellele leiti, et olenemata nädalapäevast on 80% kasutajatest vaid 10% ajast mujal kui tööl või kodus. Mobiilsusparameetrite puhul ei täheldatud erinevust tegevusraadiuses kõnetoimingute ja kogu andmestiku vahel. Tegevusraadiuse viga oli 80%-l kasutajate puhul vähem kui 1 km. Ruumientroopia puhul täheldati, et võib tekkida kallutatuse entroopia ülehindamise näol, kui kasutatakse ainult aktiivseid kasutajaid. Töö vaatlleb samuti valimi probleemi: kuidas kõnetoimingutest tingitud valim kallutab tulemusi. Saadud tulemustest järeldatakse, et kasutajatel, kelle ruumiline paiknemine on mitmekesine ning jaotunud paljude kohtade vahel, on kõnetoimingud teatud kohtade suhtes kallutatud aktiivsusega. Lisaks leitakse, et juhusliku kuid sama aktiivsusega asukohtade valim annab palju paremaid tulemusi nii tegevusraadiuse, kuid veelgi paremini ruumientroopia kohta. (Ranjan et al., 2012)

Hoteit et al. (2014) kasutasid sarnaselt Ranjan et al. (2012) oma töös kõnesid, SMS-e ning andmesidet, et valideerida liikumise trajektooride moodustamise meetodeid ning ennustada, millised mobiilimastid võiksid seetõttu saada liigset ülekoormatud.

Zhao et al (2016) võrdlesid oma töös passiivseid mobiilpositsioneerimise andmeid kogu andmestikuga, mis koguti mobiilimasti ja telefoni vahelisel suhtlusel. Töö eesmärgiks oli leida, kui täpsed on passiivsete mobiilpositsioneerimise andmetest leitud mobiilsusparameetrid. Tulemuse saamiseks jagati kasutajad 4 gruppi vastavalt kõnetoimingute ja kogu andmestiku suhtele. Tööst on võimalik järeldada, et mobiiltelefonide kasutamisest sõltuv passiivsete mobiilpositsioneerimise andmestik kallutab sellest saadud tulemusi. Kõige enam mõjutab see ruumientroopiat ning liikumisdistantsi, mis mõlemad on alahinnatud. Kõige enam mõjutab mobiilsusparameetrite kallutatust ebaühtlane telefoni kasutamine. (Zhao et al., 2016)

Eelpoolnimetatule lisaks on uuritud, kuidas ajaliselt hõredalt valimisse võetud mobiiltelefoni asukoha andmestik mõjutab inimese mobiilsuse hindamist (Lu et al., 2017). Selleks kasutati sisenevaid ja väljuvad kõnesid ning SMS-e. Samuti uuriti veel ka regulaarset asukoha uuendust, kui kasutaja vahetas mobiilimasti ning perioodilist asukoha uuendust, kui teatud aja jooksul polnud ühtegi aktiivset suhtlust mobiilimastiga toimunud. Andmed koguti Shenzheni piirkonnas ühe tööpäeva jooksul. Kokku saadi üle 16 miljoni kasutaja andmed ühelt piirkonnast

kõige suurema turuosaga mobiilioperaatori käest. Kogu andmestikust kasutati 5 miljoni kasutaja andmeid, kellel oli olemas asukohainfo igas tunnis.

1.1.1.2 Individaalsete mobiilsusparameetrite valideerimine teiste andmekogudega

Kõige varasemad tööd, mis kasutasid välimist hindamist, olid valideeritavad ja väikesemahulised. Kasutatud on vaid 37 vabatahtlikku, et tuvastada olulisi asukohti kõnetoimingute andmetest ning valideeritud neid vabatahtlike poolt kirjeldatud oluliste asukohtadega (Isaacman et al., 2011). Kasutatud on ka 38 vabatahtliku andmeid, kes kasutasid pidevaid GPS ja GSM andmeid lisaks kõnetoimingutele, et hinnata kõnetoimingute andmestiku kasutatavust, et leida olulisi kohtasid ning mobiilsusparameetreid (Schulz et al., 2012). Eeltoodud uuringus on kasutatud andmeid, mis koguti Šveitsis Lausannes ühe aasta jooksul (Laurila et al., 2013) „Mobile Data Challenge“ projekti raames.

Lisaks mobiilsusparameetritele on võrreldud ka lähte- ja sihtkohtade korrespondentsmaatriksit, passiivse mobiilpositsioneerimise andmeid koos transpordi infoga, mis koguti taksode ja busside GPS andmetest ning metroo piletisüsteemist (Zhang et al., 2014). Passiivsete mobiilpositsioneerimise andmetest saadud asukoha ajalise kehtivuse piiri on valideeritud nii mobiilse andmesidega kui ka pseudokõnetoimingu andmestikuga, mis jälgib kõnetoimingute vahelise aja tõenäosust ja on genereeritud GPS andmetes. Selleks valiti igale punktile kõige lähemal olev mobiilimast. (Chen et al., 2017a, 2017b)

1.2 Kõnetoimingutest tekkiv kallutatus

Alates passiivsete mobiilpositsioneerimise andmete kasutuselevõtust inimeste ruumilise paiknemise hindamiseks, on kaasas käinud küsimus, kui usaldusväärsete andmetega on tegemist. Kuivõrd passiivsed mobiiltelefonide andmed pole täielik valim, siis selle moodustamine võib ja tekitab kallutatust. Mitmed uuringud on toonud välja, et vaid ühe mobiilsideoperaatori kasutamine uuringus võib viia kallutatud tulemuseni. Samal ajal on probleemiks vähem jõukamad riigid ja piirkonnad, kus mobiilide arv inimese kohta on väike ning kõikuv. Sellele lisandub sotsiaalsetest teguritest mõjutatud telefonide kasutusaktiivsus (Naboulsi et al., 2016; Wesolowski et al., 2013). Eelpoolt toodule lisakson andmestiku eripära, et ruumiline ja ajaline resolutsioon ei ole ühesugune ning on pidevas muutumises.

Mõningatel juhtudel kasutatakse kõnede arvu tihedust valimi moodustamisel, aga selline tegevus viib kallutatud tulemuseni (Ranjan et al., 2012). Samas töös on ka uuritud kõnede ebaühtlast jaotumist kohtade vahel ning leitud, et kõiki olulisi paiku on raske tuvastada ainult SMS-i või kõnede abil. Samas, kodu ja töökoha leidmine 20 tööpäeva jooksul, ei erine SMS-i, kõne ja andmeside vahel 95%-l kasutajatest. Töös vaadeldi, kas kõnetoimingute arv mõjutab kogu andmestikust (SMS, kõne ja andmeside) arvutatud ruumientroopiat. Leiti, et mediaan nihkus kõige madalama kõnede aktiivsusega kvartiili kasutajatel entroopia näitajalt 3-lt 4-jale kui vaadeldi kõige aktiivsemat kasutajate kvartiili. Kasv oli läbi kvartiilide ühtlane. See kinnitab, et valides uuringusse aktiivsemad kõnetoiminguid tegevad kasutajad, kallutab ruumientroopia saadud tulemusi (Ranjan et al., 2012). Ranjan et al. (2012) uuris tegevusraadiust erinevate kõnetoimingu aktiivsusega kasutajate kvartiilide vahel ning seal olulist erinevust ei leitud. Sarnasele tulemusele jõuti ka uuringuga, kus leiti, et kõnetoimingute aktiivsemad tegijad on ühtlasi ka oma olemuselt liikuvam ja mobiilsem osa rahvastikust (Iovan et al., 2013).

Passiivselt kogutud mobiilpositsioneerimise andmete põhjal võib leida tegelikke mobiilsusväärtusi, kuid see on sõltuv mobiilsusnäitajast (Zhao et al., 2016). Samas töös tõdeti, et kuigi liikumisdistsants ja ruumientroopia on alahinnatud, olenemata kõnetoimingute ning kogu telefoni ja mobiilimasti vahelisest suhtlusest, siis tegevusraadiuse hindamist see olulisel määral ei mõjutanud. Samuti jõudsid autorid järeldusele, et enim mõjutab telefoni ebaühtlane kasutus ajas ja ruumis, mis on tingitud kasutaja poolsest telefoni kasutusest, mobiilsusparameetrite kallutatust. (Zhao et al., 2016)

Lisaks on uuritud, kuidas mõjutab erinevaid mobiilsusparameetreid (liikumisdistsants, tegevusraadius ja ruumientroopia) erinev juhuslikult genereeritud kõnetoimingute arv kogu andmestikust (kõnetoimingud koos perioodiliste ning regulaarsete asukoha uuendustega) (Lu et al., 2017).

2. Andmed ja metoodika

2.1 Uurimisala

Käesoleva uurimustöö uurimisalaks on valitud Eesti. Antud töös kasutatud GNSS andmestiku punktid on valitud Eesti majandusvööndi piirides. Kasutatud on Eesti mobiilsidevõrgus kogutud passiivse mobiilpositsioneerimise andmestikku. Sellest tulenevalt on andmestikus kõnetoimingud, mis koguti kõigilt Eesti territooriumil paiknevatest mobiilsideoperaatori antennidest. Kuivõrd mobiilsidevõrgu leviala määratlemine on sõltuv mitmetest teguritest ja on ajas muutuv, siis täpne piiritlemine on raskendatud. Töös kasutatud mobiilsideoperaatori kinnitusel katab leviala 97% Eestimaa pindalast (“Telia leviala,” n.d.).

2.2 Andmestik

2.2.1 YouSense andmestik

Käesolevas töös kasutatud andmed on kogutud spetsiaalse rakenduse – YouSense abil perioodil 22.02.2017. kuni 18.03.2018 ühe kasutaja poolt. Rakendus on loodud, et koguda efektiivsemalt mobiiltelefoni kasutajate liikumist ruumis (Linnap and Rice, 2014). YouSense andmestik on kogutud vabatahtlike poolt, kellele jagati nutitefon koos YouSense rakendusega eeldusel, et seda telefoni hakati kasutama peamise telefonina. YouSense rakendus kogub infot nii asukoha kui ka telefoni parameetrite muutumise kohta. Rakendus kogub järgmisi parameetreid:

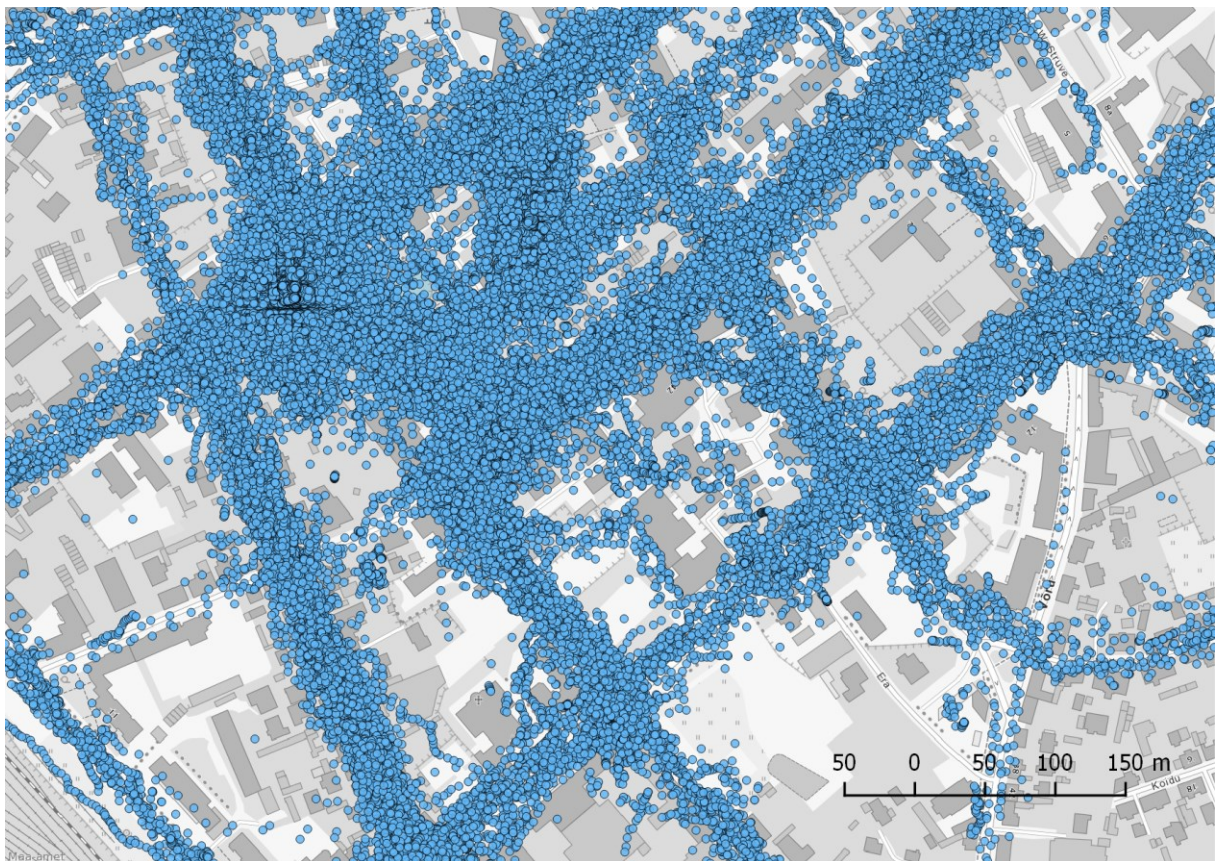
- Aeg
- Asukoht (pikkus-, laiuskraad, satelliidi aeg, täpsus, kõrgus, suund, kiirus)
- Ekraani olek
- Klahviluku olek
- Interneti ühenduse olek
- Aku laadimise olek
- Aku täituvuse %
- Lennurežiimi olek
- Äratuskella olek
- Kella ja ajavööndi seadmine
- Telefoni taaskäivitamine
- YouSense rakenduse süsteemiinfo
- Mobiilside võrguinfo

YouSense võimaldab koguda GNSS andmeid muutuva intervalliga (Joonis 1). Intervall sõltub liikumise kiirusest. Alla 3m/s liikumisel on intervall 16 sekundit ja kiiremal liikumisel 1 sekund. Intervall võib olla ka pikem kui liikumist ei toimu. Pärast liikuma asumist ei alustata andmete kogumist kohe, vaid 30-60 sekundilise viitega. Iga 10 sekundi tagant kontrollitakse

kiirendusmõõturitega vibratsiooni. Juhul kui see kordub kolmel järjestikkusel korral, siis kontrollitakse wifi võrkude pääsupunkte enne GNSSi väljalülitamist kuuldavate pääsupunktidega. Kui pääsupunktid erinevad piisavalt, siis lülitatakse wifi välja ning GNSS uuesti sisse. Selline andmete kogumise viis säästab mobiiltelefoni energiat, mis on üheks peamiseks piiravaks ressursiks pikaajaliste GNSS andmete kogumisel. GNSS andmete kogumisel tekib paus kolmel juhul:

- telefon lülitatakse välja või saab aku tühjaks;
- YouSense rakendus on pandud käsitsi pausile;
- YouSense rakendus on kasutaja poolt androidi seadetest kinni pandud.

Esimesel kahel juhul salvestatakse sellekohane kirje andmelogisse, aga viimasel puhul mitte (Vent, 2014; “YouSense,” n.d.).



Joonis 1: YouSense-i GNSS andmed Tartu kesklinnas

2.2.1.1 Yousense mobiilside võrguinfo

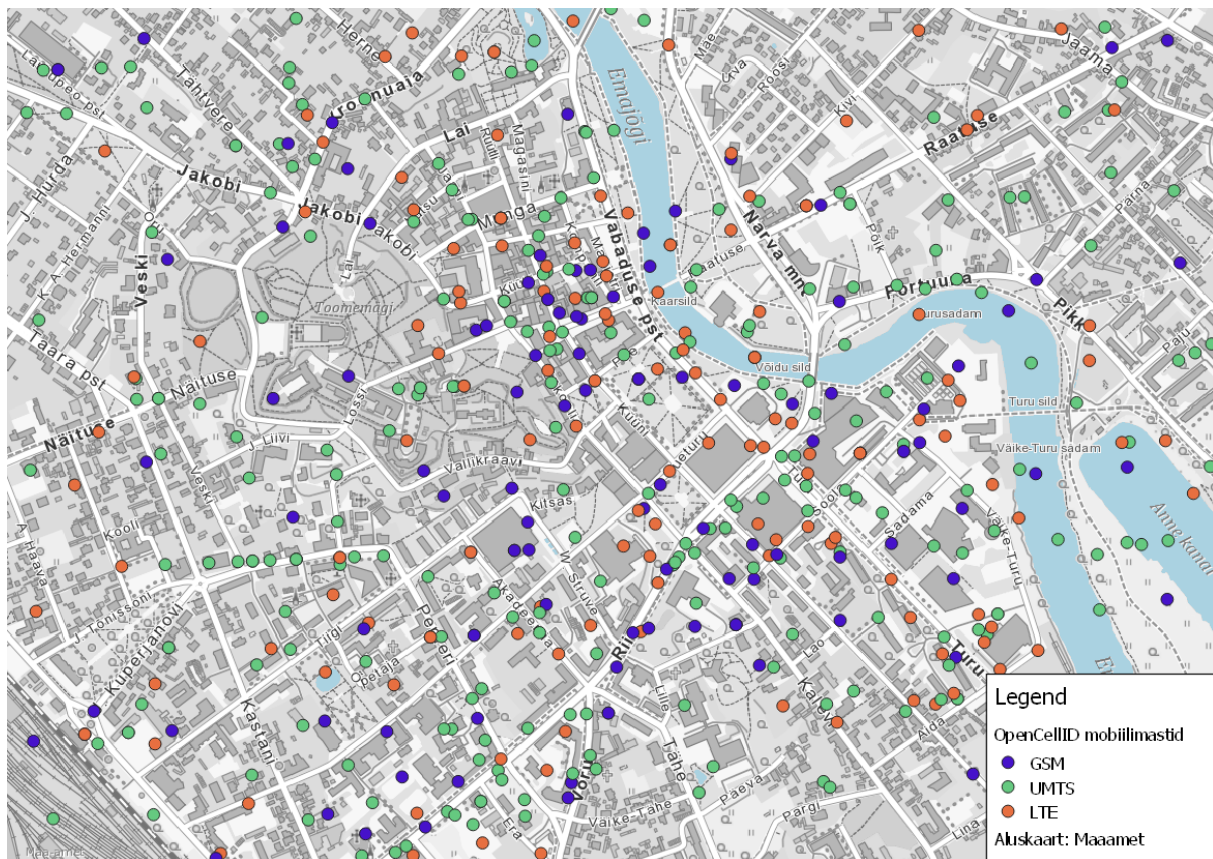
Töös kasutati YouSensi rakenduse poolt kogutud mobiilside võrguinfot. Infot võrgu kohta logiti igal korral, kui mobiilimast muutus. Iga võrgutüübi kohta (GSM, CDMA ja teadmata) olid erinevad lisaväljad, mida logiti. GSM puhul logiti järgmisi väljasid:

- võrgutüüp;
- mobiilimasti unikaalne number;
- asukoha kood;
- esmane šifreerimiskood;
- operaatori number, mis koosneb riigi ja operaatori koodist.

CDMA tehnoloogial põhinevad võrgud on praeguseks hetkeks Eestis suletud.

2.2.2 OpenCelliD mobiilivõrgu andmestik

OpenCelliD on avatud koostööprojekt, mis kogub mobiilimastide asukohaandmeid GNSS abil. OpenCelliD andmeid on võimalik kasutada Creative Commons-i litsentsi alusel. Andmestik katab kogu maakera ning see sisaldab üle 36 miljoni unikaalse GSM-i masti asukohta. Andmeid uuendab üle 75 tuhande kasutaja. Eesti andmestik sisaldab üle 36 tuhande mobiilimasti kõigilt operaatoritelt ning sisaldab GSM, UMTS, LTE mobiilivõrke (Joonis 2). Masti kohta on olemas tema asukoha koordinaadid, võrgutüüp, mobiilimasti unikaalne number, asukoha kood, riigikood ja operaatori kood. Lisaks sellele veel esmaloomise aeg ning viimase uuenduse aeg, mõõtmiste arv ja keskmine signaalitugevus ning signaali ulatus meetrites (“OpenCelliD,” n.d.).



Joonis 2: OpenCellID kõigi mobiilioperaatorite mobiilimastid vastavalt mobiilsidetechnoloogia standardile Tartu kesklinnas

3. Teekondade ja peatuspaikade eraldamine YouSense andmetest

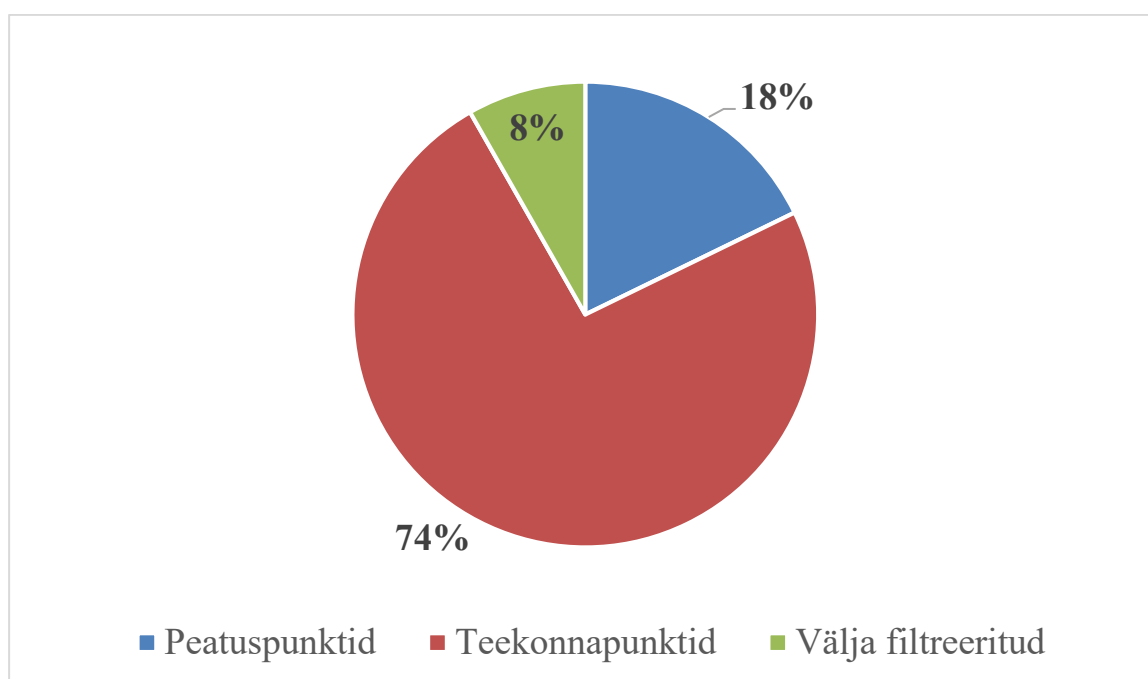
3.1.1 Andmete puhastamine

YouSense GNSS andmed puhastati vigastest andmetest, kasutades juba varasemalt väljatöötatud metoodikat (Sun et al., 2017). Käesolevas töös on kasutatud andmeid, mille täpsus oli alla 100m ja kõrgus varieerus -40m kuni 400m, mis on kohandatud Eesti tingimustega. Andmepunkt jäeti analüüsist välja, kui sellel puudusid asukoha koordinaadid, aeg või kiirus. Lisaks sellele jäeti andmepunkt valmist välja, kui selle punkti kiirus arvutatult oli GNSS seadmest üle 150km/h ning nende absoluutne vahe oli suurem kui 50km/h.

3.1.2 Peatuspaikade tuvastamine YouSense GNSS andmetelt

Asukoha punktide jagamiseks peatuspaigaks ja teekonnaks kasutati GPS andmetel põhinevate uuringute jaoks väljatöötatud peatuspaikade tuvastamise algoritmi (Cich et al., 2015; Schulz

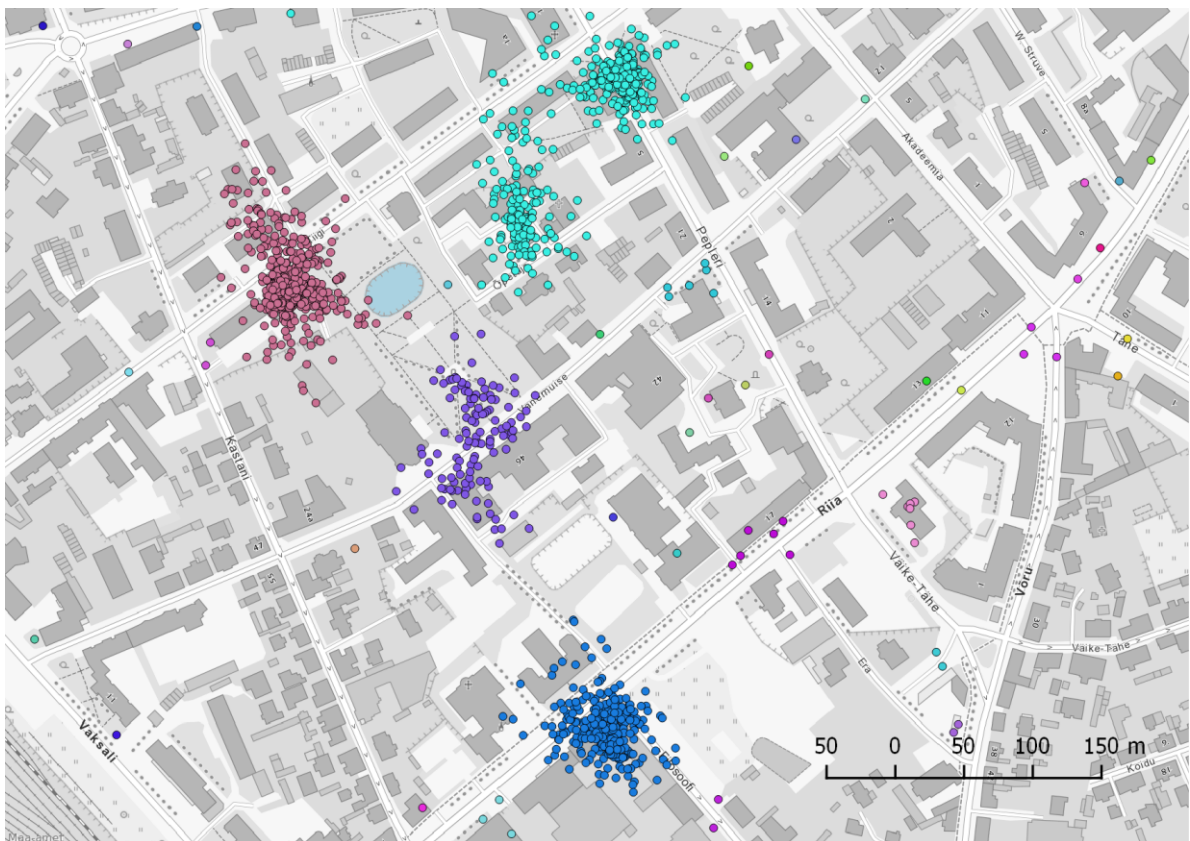
et al., 2012). Punktid järjestati ajalisel järjekorras, kasutades selleks süsteemi aega. Algoritmi parameetrite valimisel jälgiti varasemates töodes soovitud parameetreid, milleks olid peatuspaiga tuvastamist piiravad tegurid. Peatuspaigaks loeti punkti siis, kui punktide vaheline kaugus eelmiste keskmisest ei ületanud 50m ja peatuse kestvus oli vähemalt 180 sekundit ehk 3 minutit. Igale valideeritud peatuspaiga grupile omistati unikaalne identifikaator. Ülejäänud punktid, mida ei tuvastatud peatuspaigana, loeti teekonna punktideks. Joonis 33 on näha peatuspunktide, teekonnapunktide ja välja filtreeritud punktide jaotus. Kuna YouSense algoritm on kohandatud töötama energiasäästlikult ning ajaliselt tihedama resolutsiooniga liikumise ajal, siis peatuspaikade kohta on andmestik ebaühtlane ning kohati puudulik ning esineb andmeauke.



Joonis 3: YouSense GNSS andmepunktide jaotumine peatuspunktideks, teekonnapunktideks ja välja filtreerituteks.

3.1.3 Peatuspaikade grupeerimine

Eelnevalt tuvastatud peatuspaikade grupeerimiseks kasutati DBSCAN-i tihedusel põhinevat grupeerimise algoritmi (Schulz et al., 2012). Parameetrid valiti varasemate uuringute ning empiirilise põhjal. Miinimumpunktide arv 2 ja maksimaalseks kauguslõveks peatuspaikade punktide keskpunktide vahel 24m (Joonis 4). Seejärel leiti kõigi ühes grupeeritud peatuspaigas olevatele GNSS punktide keskpunkt ning omistati see ruumikuju kõigile punktidele (Joonis 5).



Joonis 4: Tuvastatud peatuspaikade grupid Tartu kesklinnas



Joonis 5: Tuvastatud peatuspaikade grupide keskmised koos ajalise kestvusega

3.2 YouSense mobiilpositsioneerimise andmete ettevalmistus

YouSense-i saadud mobiilivõrgu info ühendati OpenCellID andmekogust saadud ruumiinfoga vastavalt mobiilimasti unikaalsele numbrile, asukohale ning riigi ja operaatori koodile. Igale võrguinfo andmereale arvutati ajaline vahe, distant, kiirus ja pöördenurk ajaliselt järgmise kirjega. Seejärel jäeti välja andmerek, mille kiirus järgmisesse punkti oli üle 200 km/h ning mille pöördenurk oli 180°. Sellega välistati pingpongi efekt kahe masti vahel, mis ühendub vaid korra teise mobiilimastiga ilma, et kasutaja asukoht oleks sealjuures tegelikkuses muutunud (Iovan et al., 2013). Lisaks sellele grupeeriti kokku kõik järjestikused liikumised kahe masti vahel, sarnane liikumine nagu pingpongi efekt, kuid pikaajalisem. Leitud grupi kõigi liikmete asukoht nihutati grupeeringu kõige sagedamini esinenud mobiilimasti külge. Eelpool kirjeldatud tegevus aitab vähendada arvutatava liikumise parameetrite ebatäpsust, kuid aitab sealjuures säilitada andmete ajalise detailsuse.

3.3 Mobiilsusparameetrite arvutamine

Mobiilsusparameetrid arvutati iga andmekogu ja päeva või nädala kohta eraldi. Parameetrid arvutati samade valemite põhjal. Ainsaks erinevuseks on GPS punktidest saadud peatuspaigad, mida käsitleti esimesest kuni viimase punktini pidevana ning ajaline entroopia arvutati vaid mobiilpositsioneerimise andmetele. Päevasele, nädalasele GNSS ja mobiilpositsioneerimise andmetele arvutati kokku ajaline kattuvus ja punktide arv. Ajaline kattuvus leiti sekundites kõigi punktide vahel, mille ajaline vahe ei ületanud 15 minutit, ning näitab kui suur osa ajast oli valitud perioodil asukoht punktidega kaetud.

3.3.1 Liikumisdistant

Toetudes varasematele töödele (González et al., 2008; Lu et al., 2017; Schulz et al., 2012; Zhao et al., 2016) arvutati distant ajaliselt järjestikuste punktide eukleidilise vahemaade summana, päeva või nädala esimesest punktist kuni viimase punktini.

3.3.2 Tegevusraadius

Tegevusraadius on üks enim kasutatud tegevusruumi kirjeldamise parameetreid. Tegemist on standardhõlbega punkti ja punktikogumi keskmise vahel (González et al., 2008; Hoteit et al., 2014; Ranjan et al., 2012; Schulz et al., 2012; Zhao et al., 2016).

$$R_g(t) = \sqrt{\frac{1}{\sum W(t)} \sum_{i=1}^W W(p_i - p_c)^2}$$

$$p_c = \frac{1}{\sum W(t)} \sum_{i=1}^N W p_i$$

W on ajaline kaal, mis on arvatud igale punktile kuna ajaline resolutsioon on varieeruv. Tähis t tähistab vaadeldavat perioodi ning p_c tähistab punktide kaalutud keskmist.

Ajaline kaal leiti eeldusel, et iga punkt .Juhul kui punktide vaheline aeg on vähem kui 15 minutit, siis jagatakse aeg võrdselt mõlema punkti vahel. Aga kui kahe järjestikuse punkti ajaline vahe on rohkem kui 15 minutit, siis lisatakse mõlemale punktile ajaline kestvus 7,5 min. Juhul kui järjestikused punktid on eri päevadel ja vähem kui 15 minuti vahega, siis jagatakse aeg kahe punkti vahel vastavalt kuupäeva vahetusele.

3.3.3 Ruumientroopia

Ruumientroopiat on kasutatud andmete ühtluse määramiseks ning on fundamentaalne ennustatava kirjeldamise parameeter, mis põhineb erinevates asukohtades paiknemise tõenäosusel. Parameeter põhineb Shannoni entroopial (Lu et al., 2017; Ranjan et al., 2012; Song et al., 2010; Yuan et al., 2012; Zhao et al., 2016).

$$S_r = - \sum_{i=1}^n P_i \log_2 p_i$$

P_i näitab tõenäosust, et asukoht i on külastatud. Entroopia kasvab, mida ühtlasemalt on erinevate asukohtade vahel tõenäosused jaotunud. Skaala väikseim väärtus on 1 ning see näitab, et asukohad on jaotunud vaid ühte asukohta.

3.3.4 Ajaline entroopia

Ajaline entroopia põhineb Shannoni entroopial. See näitab andmepunktide jaotumise ühtlust.

$$S_a = - \sum_{t=1}^n P_t \log_2 p_t$$

Kus P_t näitab tõenäosust, et antud ajavahemikul t on kogutud andmepunkt. S_a on jagatud n võrdseteks vahemikeks. Töös on kasutatud 2 erinevat ajalist sammu: päev, mis on jagatud 96 võrdseks 15 minutiliseks osaks ning nädal, mis on jagatud 168 tunniks.

3.4 Korrelatsioon ja regressioonanalüüs

Käesolevas töös on arvatud kõigi parameetrite vahel Spearmani korrelatsioon ja olulisus nädalase ja päevase sammu korral eraldi. Spearmani korrelatsiooni on kasutatud, sest arvatud tunnused ei olnud normaaljaotusega ning Peasoni korrelatsioon on tundlik ekstreemumite suhtes.

Regressioonanalüüsi üks tingimusi on, et vaadeldav tunnus on normaaljaotusega. Normaaljaotuse saavutamiseks logaritmiti GNSS andmetest arvatud teekonnapiikkus, tegevusraadius, punktide arv ning mobiilpositsioneerimise andmetest arvatud teekonna pikkus, tegevusraadius, punktide arv ja ajaline kattuvus. Tulemuseks on saadud normaaljaotuse lähedased tunnused. Järgnevalt on uuritud korrelatsiooni abil GNSS andmestikust ja mobiilpositsioneerimise andmestikust arvatud parameetreid paari kaupa. Juhul kui korrelatsioon oli usaldusväärne, siis koostati normaliseeritud andmete põhjal lineaarne regressioonimudel. Kui oli teada, et regressiooni mudel läbib 0;0 punkti, siis vabaliiget mudelis ei kasutatud. Saadud regressioonimudelid on teisendatud vajadusel koos andmetega taas tagasi normaliseerimise eelsele kujule. Eelnenud tegevustele lisaks on uuritud prognoosjääkide ja mobiilpositsioneerimise andmetest arvatud ajalise entroopia, punktide arvu ning ajalise kattuvuse seost, kasutades selleks Spearmani korrelatsiooni.

4. Tulemused ja arutelu

4.1 Andmevead algandmetes

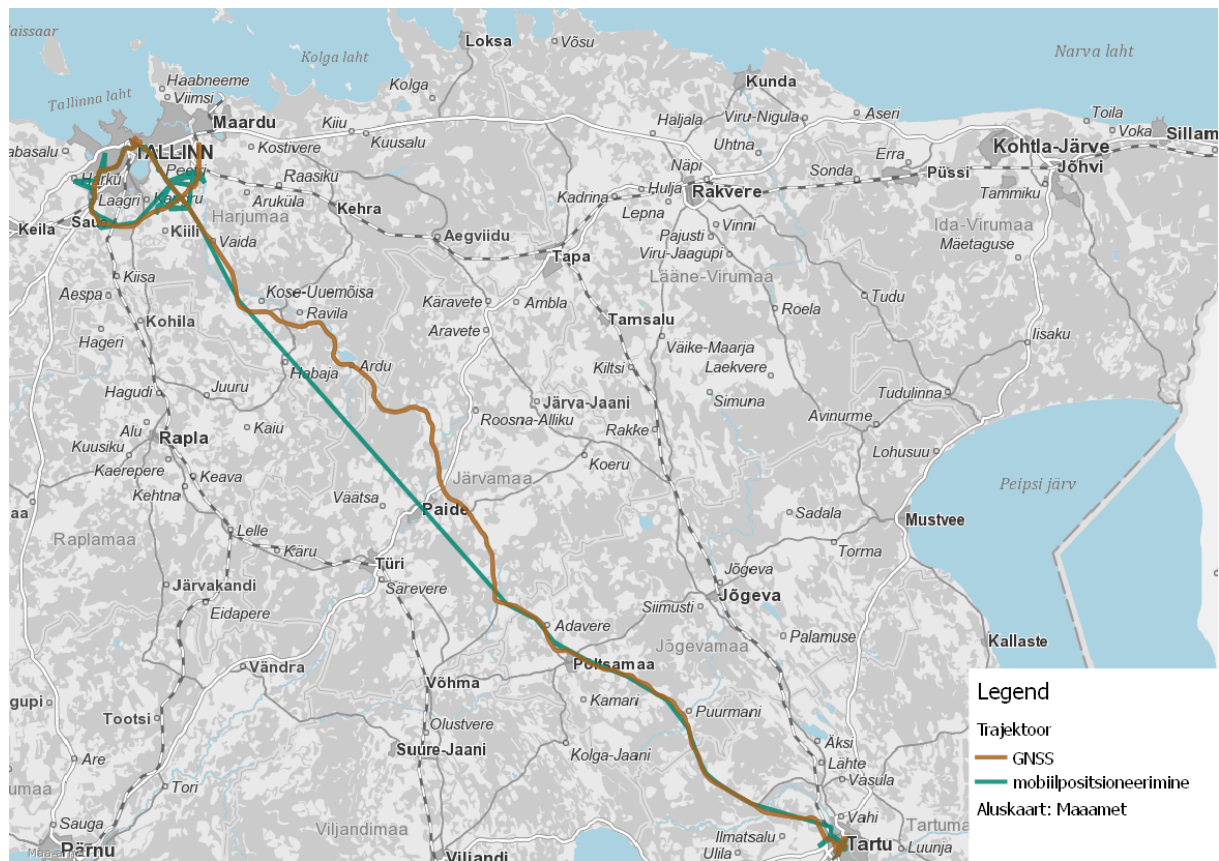
Antud töös on GNSS andmed võetud tõseks alusandmestikuks, kuid ka siin leidub vigu. Kõige olulisemaks veaks on koordinaatide triivimine tõsest asukohast eemale. Koordinaatide triivimine on tingitud seadmele kättesaadavate satelliitide arvust ning selle kvaliteedist, mida mõjutavad ümbritsevad keskkonna tingimused. Üheks peamiseks probleemiks on signaali peegeldumine ja murdumine keskkonnas, mis põhjustab omakorda signaali teekonna pikendamise. Viga tuleb esile siseruumides, kus erinevate satelliitide signaale on vähe ning kohale jõudvad signaalid on suure tõenäosusega jõudnud kohale läbides pikema teekonna. Seetõttu on siseruumides saadud andmed ebatäpsemad. Sellele lisaks mõjutab GNSS andmete täpsust kanjoni efekt. See tähendab, et seade on paljudest külgedest ümbritsetud signaali mitteläbilaskvate kehade ja nähtav on vaid väike osa taevast. Sarnane olukord tekib kanjonis ning seetõttu on seadmele näha vaid kitsas piirkonnas olevad satelliidid. Satelliitide vaheline nurk aga parandab tulemuste täpsust samas kui väga kitsa taevapiirkonna satelliitide tulemuste täpsus väheneb. Töös kasutatud GNSS filtrid suudavad enamuse ebatäpsetest asukohtadest kõrvaldada, kuid siiski mitte kõiki. Antud vea suhtes on kõige tundlikum liikumisdistantsi parameeter. Mida väiksem on liikumisdistants, seda tundlikum parameeter on. Kuivõrd peatuspaigad asuvad enamuses siseruumides, siis on peatuspaikade andmestik enam mõjutatud antud veast kui teekonna punktidest.

Teiseks veaks, mis YouSense GNSS andmestik sisaldab, on andmeaugud. YouSense kasutab energia efektiivsuse saavutamiseks ära eeldust, et kui telefon ei rapu, siis tema asukoht ka ei muutu. Sellest on tingitud pikad ajavahemikud, kus ei ole ühtegi asukohapunkti. Neid andmeauke ei ole võimalik eristada GNSS andmestikus kasutaja poolt teadlikult rakenduse pausile panemisest, GNSS seadme välja lülitamisest, seadme tühjaks saamisest või väljalülitamisest. Lisaks sellele ei toimi algoritm sihipäraselt kui seadme WIFI võrk on välja lülitatud. Seadme WIFI võrgu keelamine ei võimalda kontrollida WIFI pääsupunktide nimekirja ning seega, ei lülitata sisse ka GNSS seadet. Samuti võib liikumise alustamine jääda märkamata tänu sellele, et ei saada kätte kolme järjestikust rappumist 10 sekundilise vahega, mis käivitaks GNSS seadme. Kirjeldatud nähtus võib olla tingitud sellest, et kasutaja sõidab rongi, bussi või autoga. Juhul kui telefon on taskus ning liigutakse jalgsi, siis sellist probleemi ei tohiks esineda ("YouSense," n.d.).

Lisaks kirjeldatakse YouSense infolehel, et androidi süsteem on ebausaldusväärne ja ei raporteeri kõigist sündmustest, mida rakendus kuulab. Näiteks ei teavitata rakendust rappumise toimumisest, mis käivitaks kontrolliahela GNSS süsteemi käivitamiseks. Võib ka juhtuda, et rakendus, ei saa kätte uuendatud asukoha infot. Rakendus ei võimalda kirjade logimist kui süsteemi mälu on täis ning uusi kirjeid ei ole võimalik salvestada juurde (“YouSense,” n.d.).

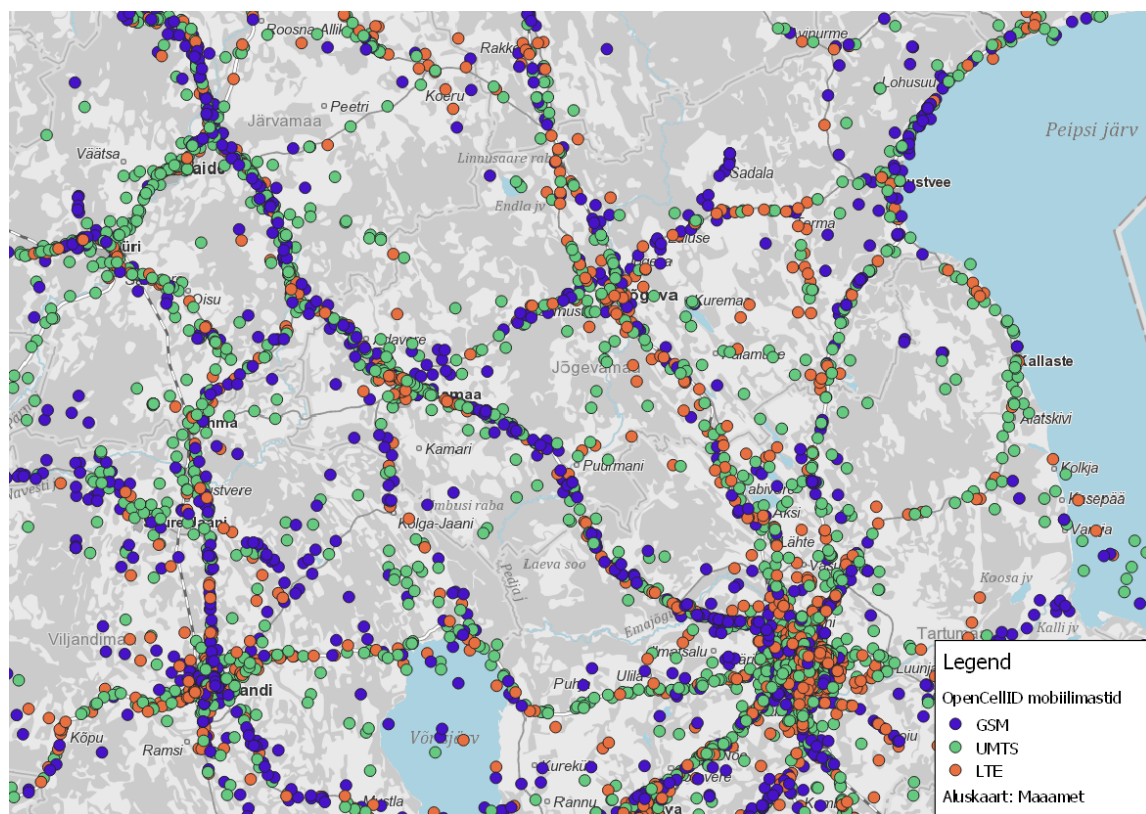
YouSense kogub ühenduses oleva mobiilivõrgu andmeid pidevalt ning salvestab logisse kui telefon seda uuendab või see muutub. Eestis on kasutusel kolme põlvkonna mobiilivõrgud: GSM, UMTS ja LTE. Kõnetoiminguid on võimalik teha kahe esimesega. Rakendus logib kõiki võrke, millega ta on ühenduses, samas 85% kogutud mobiilivõrgu andmetest on LTE võrgus, 13% UMTS ja vaid 2% GSM võrgus. Kuna kõik kolm võrku on üksteist dubleerivad, siis piisab vaid kõnetehnoloogia vahetusest, näiteks kõnetoimingu tegemisest, selleks, et tekiks tehislik liikumine. Kui varasemates töödes on kirjeldatud pingpongi efekti (Iovan et al., 2013), mis on lühiajaline. YouSensi poolt kogutud andmed sisaldavad pikemaajalist kahe masti vahel liikumist. Pikemaajaline kahe masti vahel liikumine on tingitud osaliselt sellest, et telefon valib parima võrgutehnoloogia ja parima signaaliga masti. Juhul kui telefon asub kahe masti vahel ja signaali tugevused on võrdsed, siis keskkonnatingimuste muutudes signaali tugevus varieerub. See omakorda põhjustab masti vahetust ning tehislikku liikumist. Peatükis 3.2 on välja pakud meetod mobiilpositsioneerimise andmete puhastamiseks.

YouSensi mobiilpositsioneerimise andmekogu on osaliselt puudulik. Varasemalt mainitud androidi operatsioonisüsteem on ebausaldusväärne ning mõned sündmused ei jõua rakenduseni. Arvatavasti on see üheks põhjuseks, miks on mobiilpositsioneerimise andmestikus auke, kus pole ühtegi masti vahetust registreeritud rakenduse poolt, samas on GNSS punkte kogutud ning asukoht on muutunud piisavalt. Üheks näiteks on (Joonis 6) teekond Tallinnast Tartusse, kus YouSense rakendus on kogunud GNSS andmeid pidevalt, kuid mobiilivõrgu andmestik puudub Kosest Imavereni ning peale seda taastub mobiilivõrgu info kogumine. Vaadeldud andmeaukude täielik leidmine andmestikust on keeruline kuna väiksematel distantidel võibki mobiiltelefon püsida ühe mobiilimasti levialas.



Joonis 6: YouSense mobiilpositsioneerimise andmeauk teekonnal Tallinn - Tartu

OpenCellID on kogutud vabatahtlike abil, kes koguvad isiklike mobiiltelefonide abil oma asukohtasid ning võrgu infot. Saadud andmetest on koostatud tõenäolised mobiilimastide asukohad. Tulenevalt sellest, et andmestiku kogumine on ebaühtlane ning linnadest väljaspool koondub suurematele teedele (Joonis 7), siis esineb masti tõelise asukoha ja kogutud andmetest genereeritud masti asukohtade vahel suuri vigu, mis kasvavad masti leviala suurenedes asulavälistel piirkondades. Samuti esineb OpenCellID mastide valesid asukohti (Joonis 8). Osaliselt võivad sellised vead esineda suure levialaga mastidel ning kui OpenCellID andmed on koondunud kindlasse punkti ning kogutud punkte on vähe, siis võib viga olla kuni maksimaalne leviala kaugus. pole tagatud OpenCellID kogutud andmestiku kvaliteet ning seda pole võimalik ka tagantjärele kindlaks teha. Mobiilimastide asukohad ning levialad ei ole staatilised, sest operaator kohandab seda vastavalt nõudlusele. Mobiilimastide leviala mõjutavad samuti keskkonnatingimused, sealhulgas reljeef, pinnakate, sademed.



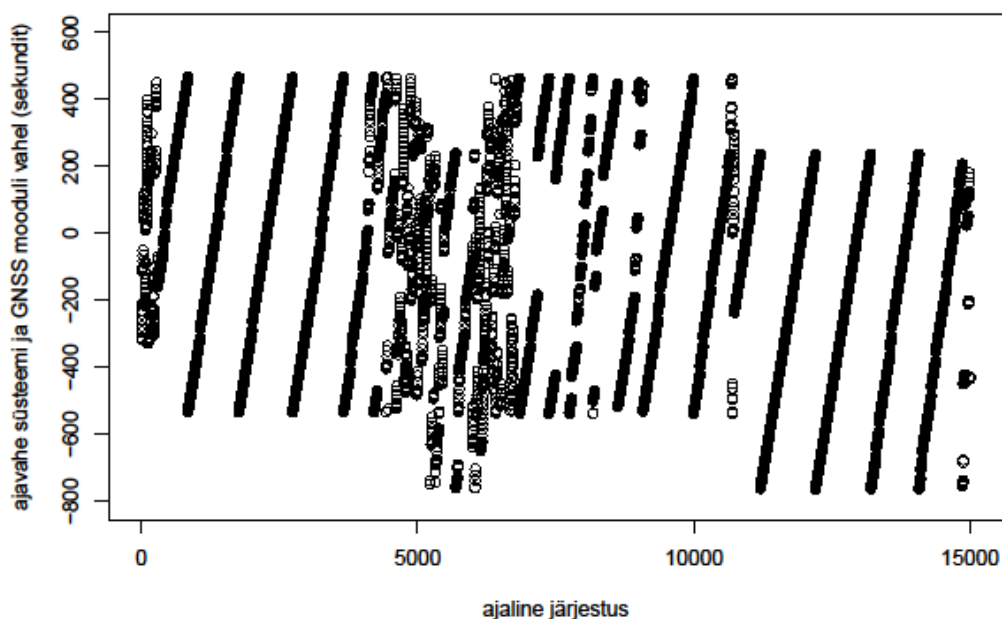
Joonis 7: OpenCellID mobiilimastide koondumine põhiteedele



Joonis 8: OpenCellID mobiilimastide vale asukoht

YouSense andmetega töötlemisel on üheks probleemseks kohaks andmete ajatempel. Rakendus logib nii süsteemi kui GNSS moodulist tulevat kellaaega UTC ajavööndis millisekundites alates 01.01.1970. Mõlemal juhul esineb vigu. Võrreldes GNSS mooduli ja süsteemi ajalist vahet ilmneb, et süsteemi aega uuendatakse mõningatel juhtudel vaid iga 1000 sekundi tagant. *Seega kord on süsteemi aeg ees ning kord taga, vastavalt andmete salvestamise intervallile jääb GNSS mooduli ajast maha* (Joonis 99). Võimalikud vead tulenevad osalt sellest, et süsteemi aeg on kasutaja ja süsteemi poolt muudetav. GNSS moodulist tulev aeg peaks olema täpne, kuid seda ei ole võimalik siduda teise sündmustega. Androidi operatsioonisüsteem jagab vahel valesid aegasid. Süsteemi aeg nullitakse kui aku on eemaldatud või tühjaks saanud. Peale taaskäivitamist pannakse vaikimisi ajaks 1980 (“YouSense,” n.d.). Lisaks kellaaegadele on igale kirjele omistatud ühe võrra suurenev numbriline loendur, mis võimaldab tagada kirjete ajalise järjestuse. Loendur on vaatlusperioodil mitmeid kordi nullitud.

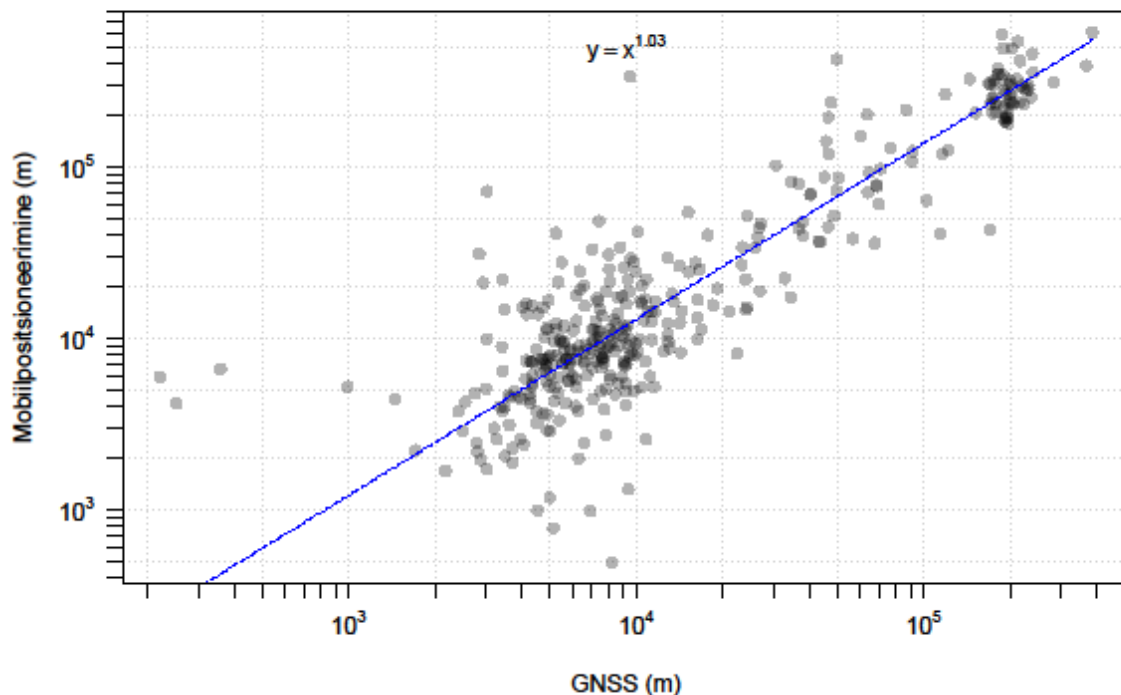
Mõlemad nii YouSense-i mobiilpositsioneerimise kui GNSS andmed sisaldavad vigu ja andmeauke, mis mõjutavad töötulemust. Töös on kasutatud GNSS andmeid tõelise andmestikuna kuid seal leiduvad vead võivad tekitada vigu tulemustes. Sellest lähtuvalt tuleks tulevastest töodes täpsemalt uurida YouSense GNSS andmete vigu ja kõrvutades neid mõne sama väärse või täpsema andmestikuga.



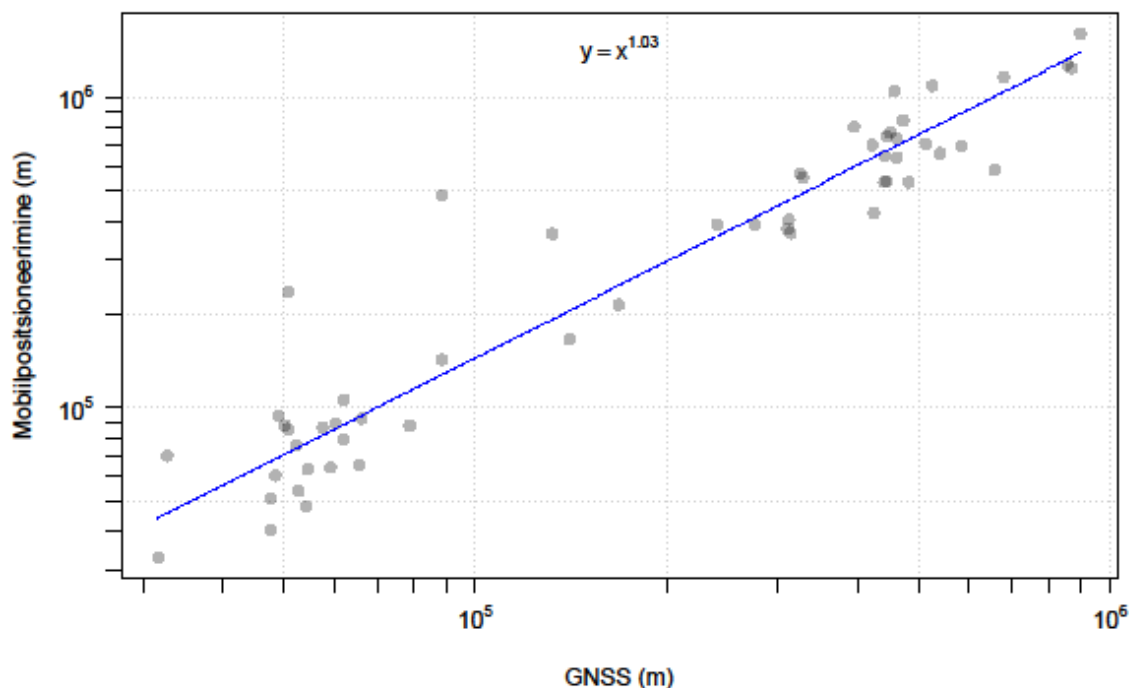
Joonis 9: Ajavahe muutus süsteemi ja GNSS mooduli aja vahel

4.2 Liikumisdistsants

Mobiilpositsioneerimise- ja GNSS andmetest arvatud liikumisdistsantside vahel on väga tugev ning statistiliselt oluline korrelatsioon. Spearmani korrelatsiooni kordaja on päevase sammuga mõõdetuna 0,79 ning nädalase sammuga 0,92. Mõlemal juhul on olulisuse näitaja väärtus väiksem kui 0,001. Regressioonanalüüsi korrelatsioonikordaja ruut näitab ära kui suure osa variatsioonist on mudel võimeline ära kirjeldama. Päevase sammu korral kirjeldab mudel 99% ning nädalase sammu puhul 100% liikumisdistsantsi variatsioonidest. rjastandardviga on vastavalt 48988 m ja 133600 m. Regressiooni mudel on $y = x^{1,04}$ päevase sammuga ning $y = x^{1,03}$, kus y on mobiilpositsioneerimise andmistest arvatud liikumisdistsants ja x on GNSS andmetest arvatud liikumisdistsants.



Joonis 10: Mobiilpositsioneerimise ja GNSS andmetelt arvatud liikumisdistsants päevase sammuga



Joonis 11: Mobiilpositsioneerimise ja GNSS andmetelt arvutatud liikumisdistsants nädalase sammuga

Mobiilpositsioneerimise andmed on väga tugevas korrelatsioonis GNSS andmetega. Kõige paremini korreleeruvad andmed liikumisdistsantsi ja tegevusraadiuse parameetrite võrdluses. Antud parameetrid on samuti omavahel tugevas seoses, sest suurem tegevusraadius suurendab ka liikumisdistsantsi. Mõlemad parameetri mudelid on astmefunktsioonid, kus astendaja on ligilähedane 1-le. Selle järgi võib kinnita, et nii üle- kui alahindamisi on võrdselt. Regressioonmudeliks valiti astmefunktsioon, sest mõlemad parameetrid tuli logaritmida, et saavutada normaaljaotus, mis on üks eeldusi regressioonanalüüsiks. Teisendatud parameetrite vahel leiti lineaarne mudel, mis tagasi transformeerimise käigus teisendus astmefunktsiooniks. Liikumisdistsantsi ja tegevusraadiuse mudelid nii päevase kui nädalase sammuga kirjeldavad ära 98%-100% andmetest, mis näitab, et tegemist on väga heade regressioonimudelitega.

Vaatamata sellele, et liikumisdistsantsi mudelid on head, esineb neis ekstreemumeid. Kõige enam tulevad need esile väikestel distantsidel, kus vigadel on suur mõju lõpptulemusele. Üheks probleemiks on juhud kui mobiilpositsioneerimise andmed alahindavad tegelikku liikumisdistsantsi. Selle põhjuseks võib olla ühe ja sama mobiilimasti piirkonnas liikumine,

mis ei kajastu mobiilmasti vahetusega. Selline viga esineb hajaasustuse aladel, kus üks mobiilimast katab suurt territooriumi. Antud andmestikus selliseid vigu esineb vähe, sest enamuse andmeid on kogutud tiheasustusaladel, kus mobiilimastide võrk on tihedam.

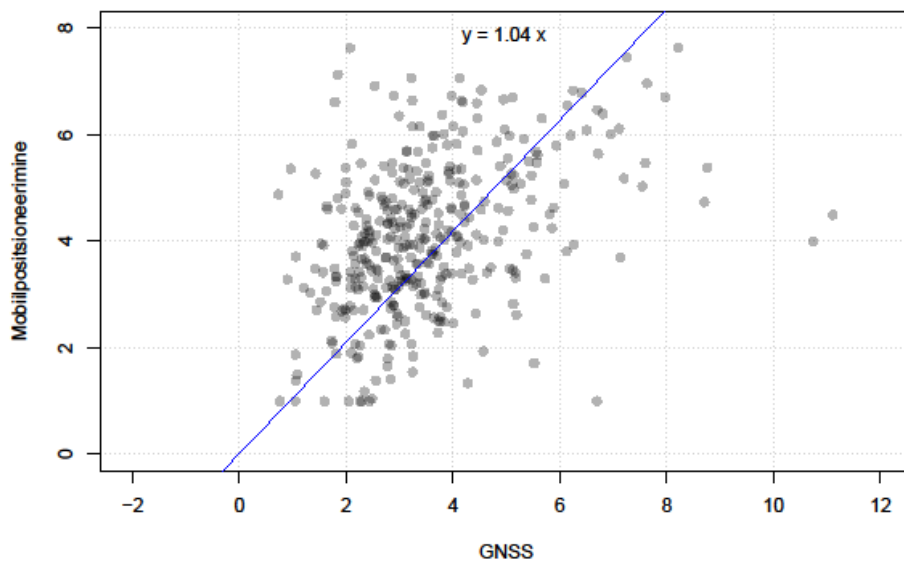
Teiseks võimalikuks probleemiks on YouSense poolt kogumata jäetud võrgu info, mille tulemusena on mobiilpositsioneerimise andmetest leitud teekonna pikkus väiksem kui tegelikult läbitud teekonna pikkus. Kui antud vea korral alahinnatakse liikumisdistsantsi, siis esineb ka liikumisdistsantsi ülehindamist. Sellisel juhul on kasutaja olnud paigal või on liikunud väiksema distantsi kui mobiilivõrgu järgi võiks eeldada. Joonis 12 järgi on kasutaja püsinud päeva jooksul ühes asukohas, kuid selle aja jooksul on mobiiltelefon olnud ühenduses mitmete erinevate mobiilimastidega kuni 2 km kaugusel. Antud näitel on väga selgelt viga eristatav. Kui antud viga toimub pidevalt ning selle tõttu on liikumisdistsants üle hinnatud. Kuivõrd kirjeldatud vead mõjutavad oluliselt rohkem väiksema distantsiga päevasid (Joonis 10), siis mudel toimib korrektselt alates 2-3km. See võib erineda olenevalt mobiilimastide tihedusest. Nädalase sammuga (Joonis 11) on vead enam silutud ning seeläbi on ka mudeli näitajad paremad.



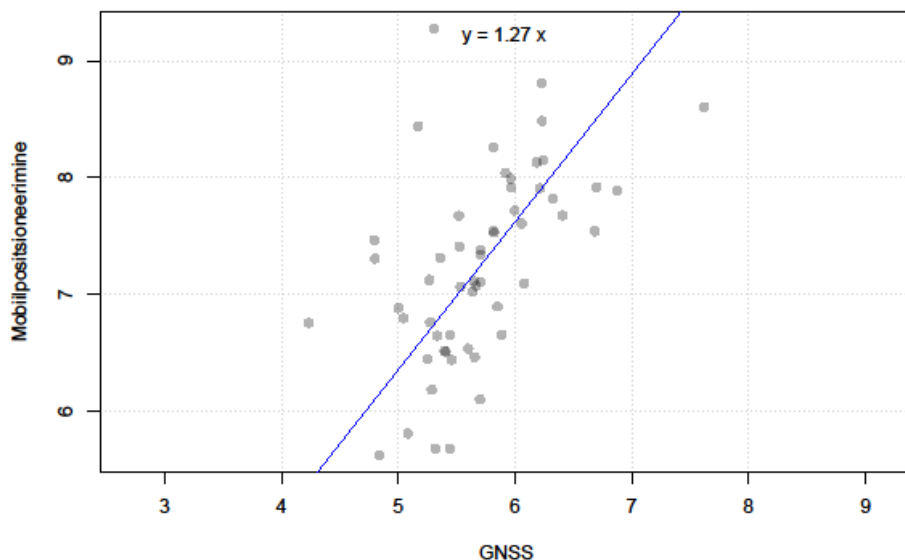
Joonis 12: Mobiilimastide vahetus põhjustab vigu mobiilsusparameetrites

4.3 Ruumientroopia

Ruumientroopia andmete vahel olev Spearmani korrelatsiooni seos on keskmise tugevusega, olles päevase sammu korral 0,41 ning nädalase sammuga 0,60. Mõlemal juhul on olulisuse näitaja p väiksem kui 0,001. Regressioonanalüüsi standardviga on 1,72 ja 0,75 ning korrelatsioonikordaja ruut on päevase sammuga 0,84 ja nädalase sammuga 0,99. Mudelid olid vastavalt $y = 1,04x$ ja $y = 1,27x$, kus y on mobiilpositsioneerimise andmetest arvatud ruumientroopia ja x on GNSS põhjal arvatud ruumientroopia (Joonis 13, Joonis 14).

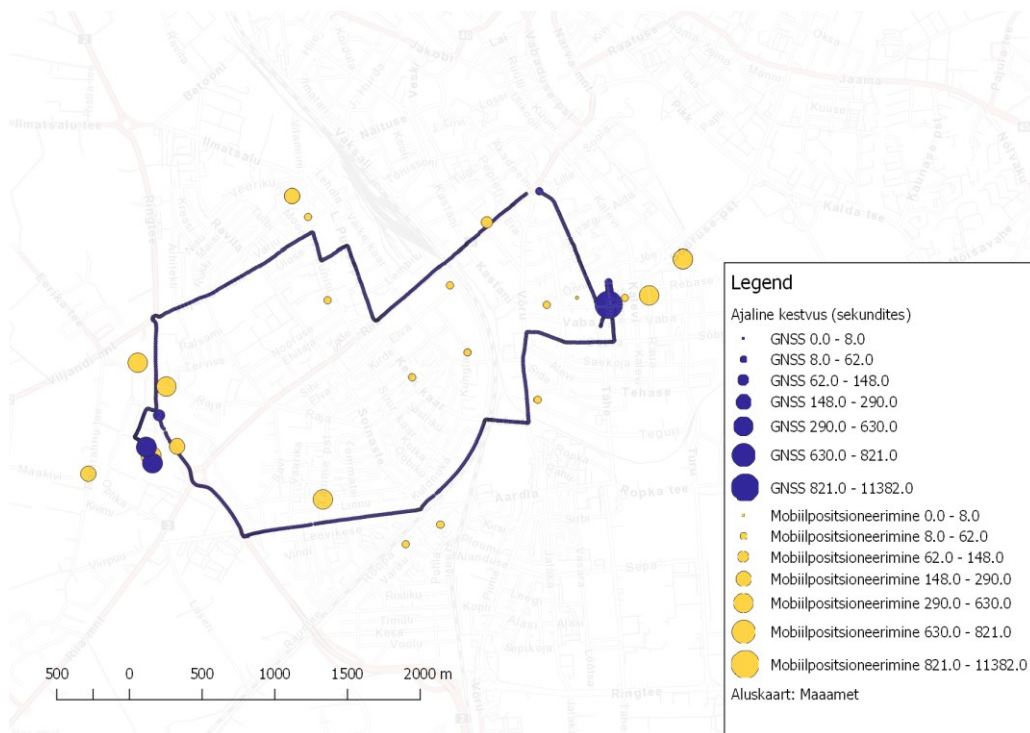


Joonis 13: Mobiilpositsioneerimise ja GNSS andmetest arvatud ruumientroopia päevase sammuga



Joonis 14: Mobiilpositsioneerimise ja GNSS andmetest arvutatud Ruumientroopia nädalase sammuga

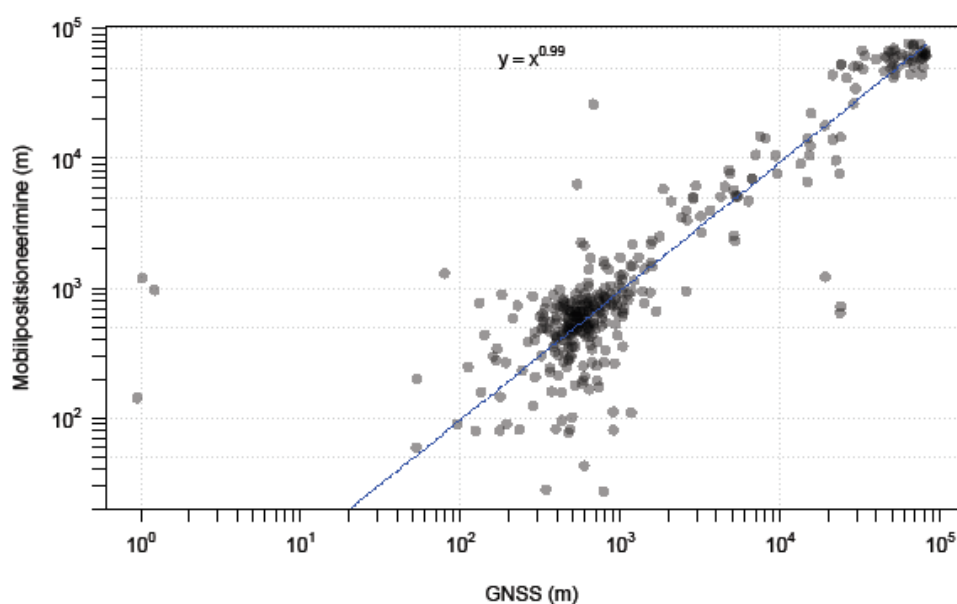
Ruumientroopia parameetrite korrelatsioon on tunduvalt nõrgem liikumisdistsantsi ja tegevusraadiuse omast. Kui päevase sammuga on nii üle- ja alahindamisi enam-vähem võrdselt ning regressioonimudeli kordaja on ligilähedane 1-le, siis nädalase sammuga on keskmiselt 1,27 korda ruumientroopia ülehinnatud. Kuigi nädalase sammuga regressioonimudel hindab üle ruumientroopiat, siis mudel kirjeldab suuremat hulka kirjetest kui nädalase sammuga. Üks ruumientroopia ülehindamise põhjustest on pidev mobiilimasti vahetus ka paigal seistes, mis hajutab mobiilpositsioneerimisest saadud asukohtasid ühtlasemalt erinevate kohtade vahel laiali ning, mille tõttu ruumientroopia kasvab (Joonis 15). Samuti on see seotud mobiilimasti võrguga, sest tiheasustusaladel on mobiilimastide levialad mitmekordselt kattuvad, mis annab suurema tõenäosuse mastivahetuseks kui piirkonnas, kus ühe mobiilimasti leviala katab ära mitmeid kasutaja jaoks olulisi paikasid. Sealjuures põhjustab suur mobiilimasti leviala ruumientroopia alahindamist kuna erinevaid asukohti saadakse vähem kätte ning see langetab ruumientroopia näitajat.



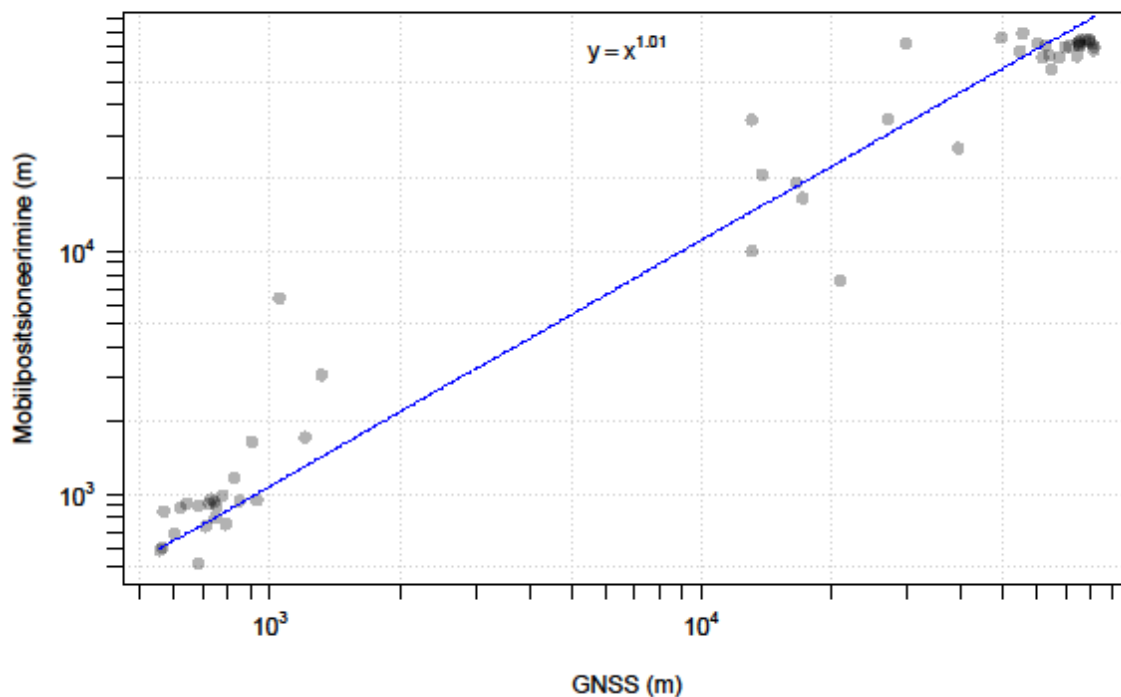
Joonis 15: GNSS ja mobiilpositsioneerimise andmete ajalis-ruumiline jaotumine ühe päeva lõikes

4.4 Tegevusraadius

Tegevusraadiuse Spearmani korrelatsioon päevase sammuga on 0,76 ning nädalase sammuga 0,91, mis mõlemad on väga tugevad seosed. Mõlema korrelatsiooni olulisuse näitaja on väiksem kui 0,001. Regressioonanalüüsi standardviga on 6560 ja 11416 meetrit. Korrelatsioonikordaja ruut 0,98 ja 1,00. Regressiooni mudel on päevase sammu korral $y = x^{0,99}$ ja nädalase sammu korral $y = x^{1,01}$ (Joonis 16, Joonis 17).



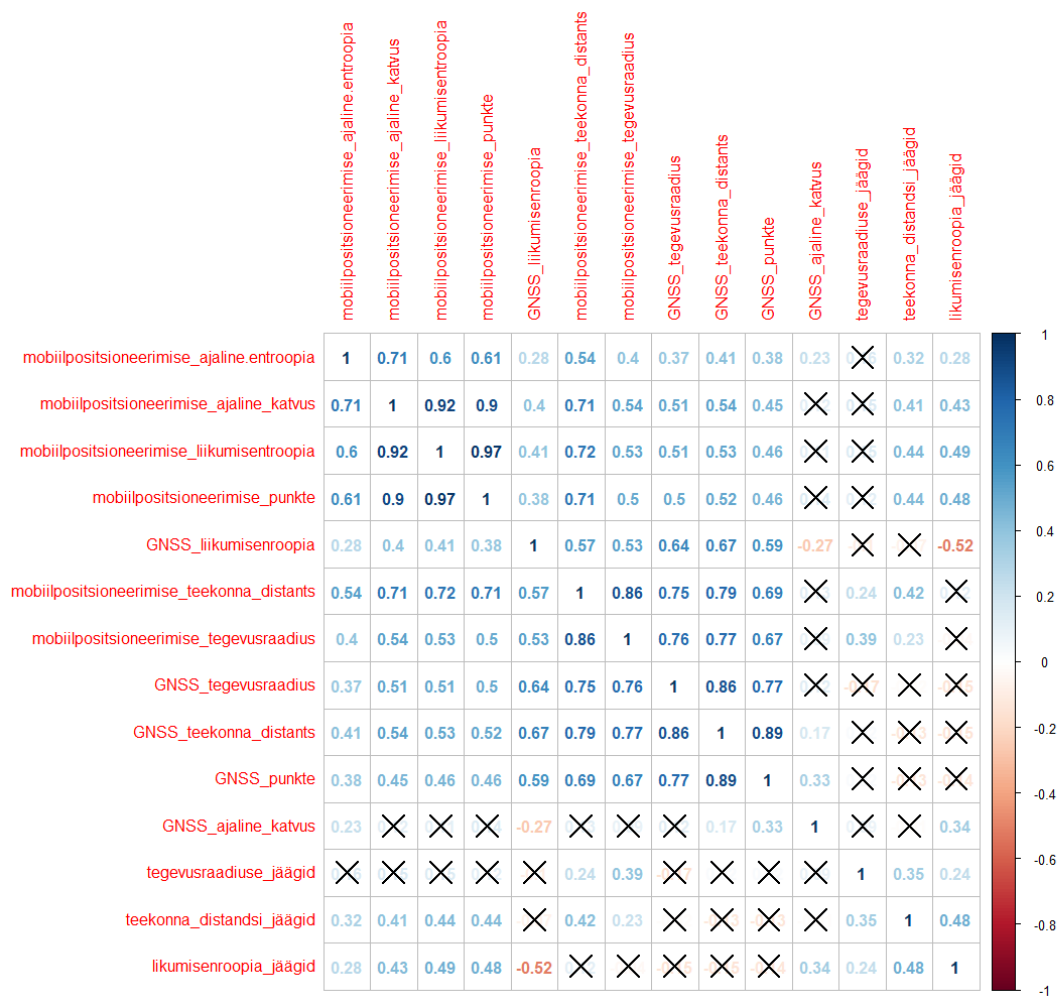
Joonis 16: Mobiilpositsioneerimise ja GNSS andmetest arvatud tegevusraadius päevase sammuga



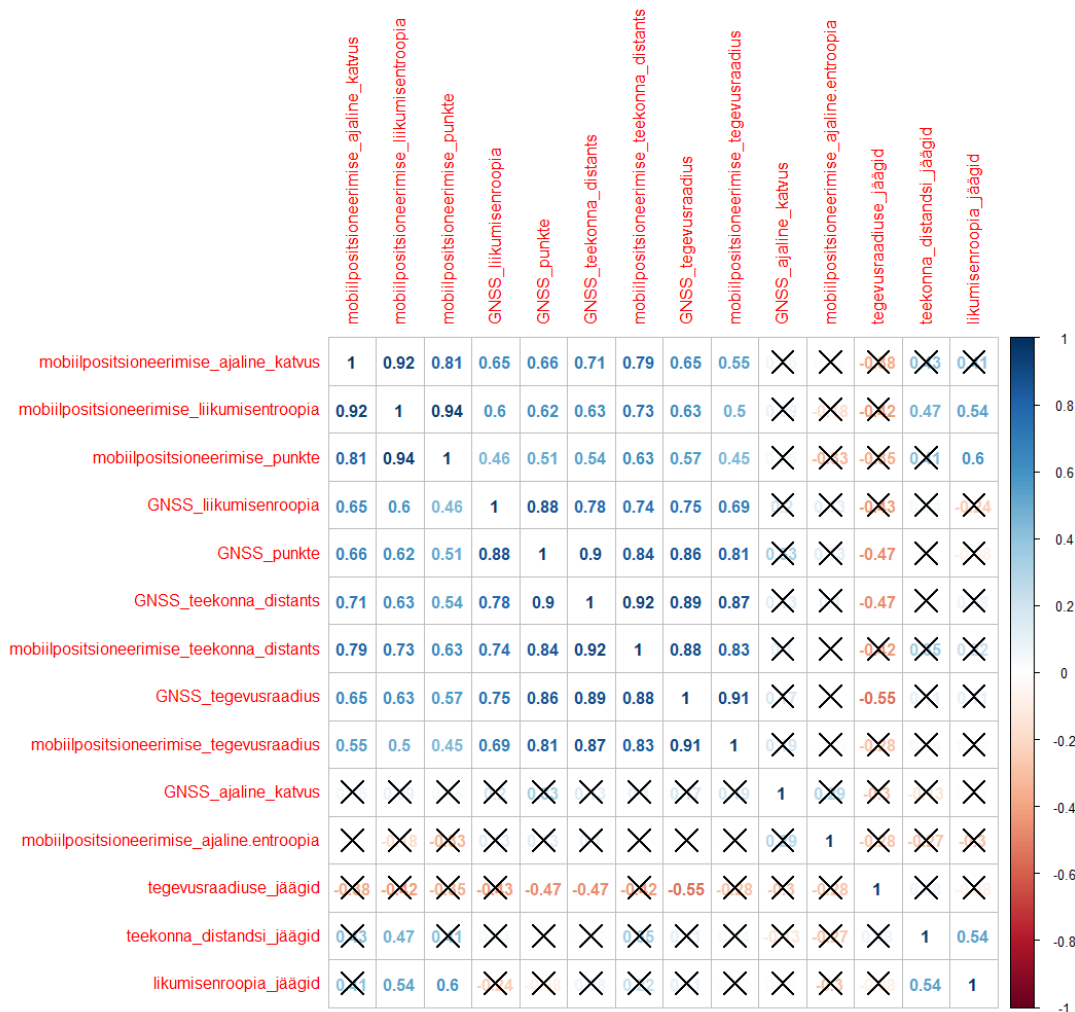
Joonis 17: Mobiilpositsioneerimise ja GNSS andmetelt arvutatud tegevusraadius nädalase sammuga

4.5 Lisaparameetrid

Kõikide arvutatud parameetrite Peasoni korrelatsiooni maatriksiga on võimalik tutvuda (Lisa 1, Lisa) vastavalt päevane ja nädalane samm. Tugev ja oluline korrelatsioon erines järgmiste muutujate vahel (Joonis 18, Joonis 19). Jooniste värvusskaala näitab kahe parameetri tugevust ning ristiga on kaetud seosed, mille olulisus oli $p > 0,001$



Joonis 18: Spearmani korrelatsioon ja olulisus parameetrite vahel päevase sammuga. Korrelatsioonid mille olulisus on $P > 0,001$ on maha tõmmatud ristiga



Joonis 19: Spearmani korrelatsioon ja olulisus parameetrite vahel nädalase sammuga
Korrelatsioonid mille olulisus on $P > 0,001$ on maha tõmmatud ristiga

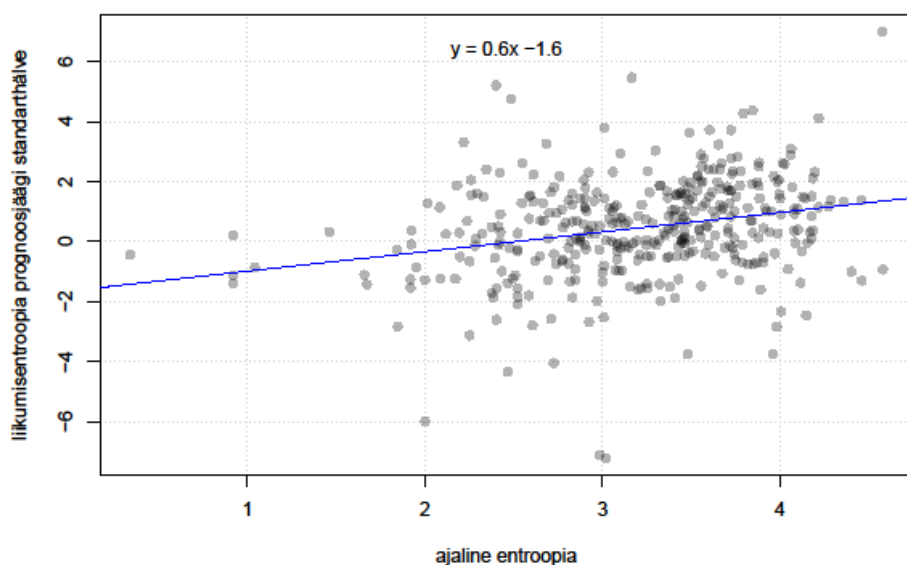
4.6 Ruumientroopia prognoosjäägi standardhälbe regressioonanalüüs

Lisaks liikumisdistsantsile, ruumientroopia ja tegevusraadiusele koostati regressioonanalüüs: ruumientroopia hälve regressiooni mudelist ning ajalise katvuse, ajalise entroopia ja mobiilpositsioneerimis punktide arvu vahel. Enne regressioonanalüüsi oli logaritmimise teel mobiilpositsioneerimise andmete ajaline kattuvus ja punktide arv viidud normaalkujule. Selleks võeti kõigist vajalikest muutujatest logaritm.

Mobiilpositsioneerimise ruumientroopia prognoosjäägi standardhälbe ja katvuse, ajalise entroopia ja GNSS ruumientroopia vahel oli vaid päevase sammu korral oluline seos. Samas mobiilpositsioneerimise punktide ja ruumientroopia seos ruumientroopia prognoosjäägi standardhålbega oli seose olulisus nii päevase kui nädalase sammuga oluline ning üle keskmise tugevus seosega (Tabel 1).

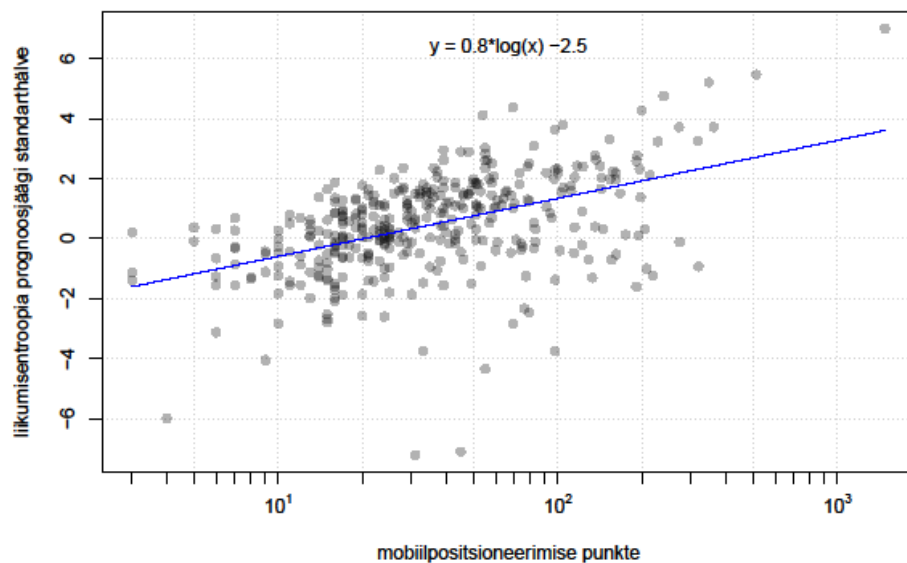
Tabel 1: Ruumientroopia prognoosjäägi standardhälbe olulisemad seosed

samm		mobiilpositsioneerimine				GNSS
	parameeter	punkte	ajaline entroopia	ruumientroopia	ajaline kattuvus	ruumientroopia
päev	ruumientroopia	0,48	0,28	0,49	0,43	-0,52
nädal	prognoosjäägi standarthälve	0,6	X	0,54	X	X

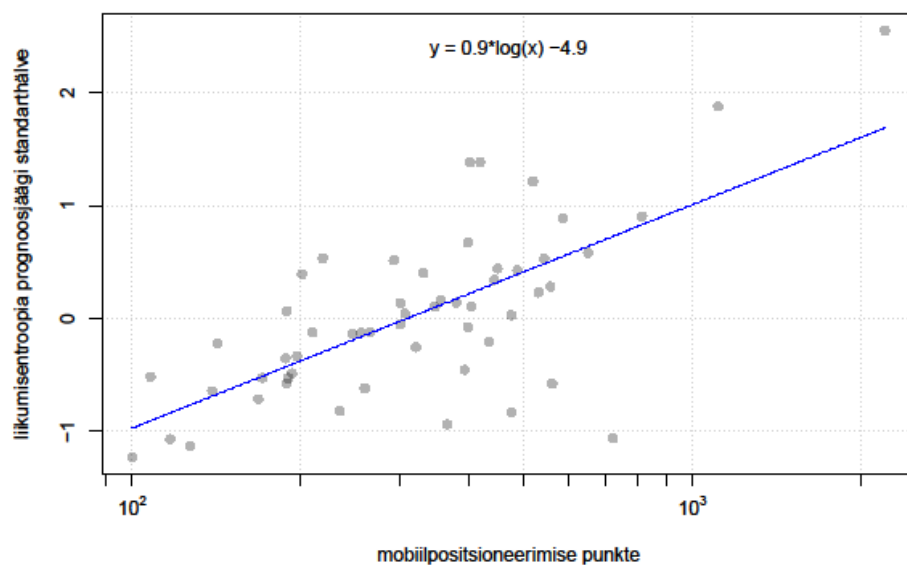


Joonis 20: Ruumientroopia prognoosjäägi standardhälbe seos ajalise entroopiaga päevase sammuga

Ruumientroopia prognoosjäägi standardhälbe Spearmani korrelatsioon ajalise entroopiaga päevase sammuga on oluline, kuid nõrk ning nädalase sammu puhul on seose olulisus suurem. Regressioonimudel (Joonis 2020) kirjeldab ära vaid 6% kogu andmetest ning jääkide standardhälve on 1,61.



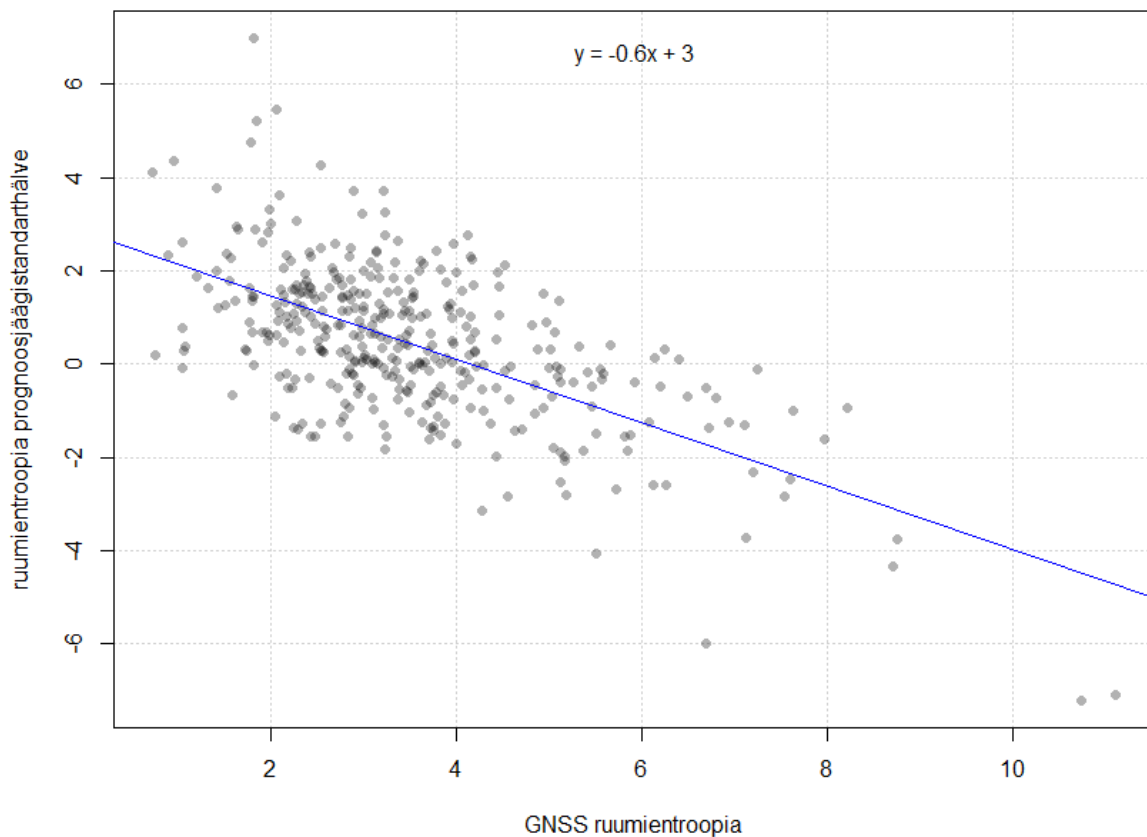
Joonis 21: Päevase sammuga ruumientroopia prognoosjäägi standardhälbe seos mobiilpositsioneerimise punktide arvuga



Joonis 22: Nädalase sammuga ruumientroopia prognoosjäägi standardhälbe seos mobiilpositsioneerimise punktide arvuga

Tugevaim seos on ruumientroopia prognoosjäägi standardhälbe ja mobiilpositsioneerimise-punktide vahel. Spearmani korrelatsioon on 0,48 ja 0,60 vastavalt päevase ja nädalase sammuga. Regressioonimudelid kirjeldavad ära päevase sammuga 22% ning nädalase sammuga 47% andmestikust. Standardhälbed on vastavalt 1,47 ja 0,55 (Joonis 21, Joonis 22).

Väga tugev negatiivne Spearmani korrelatsioon on ruumientroopia jäägi standardhälbe ja GNSS andmetest arvatud ruumientroopia vahel (Joonis 23). Spearmani korrelatsioon on -0,53 ning olulisus $p < 0,001$. Mudel kirjeldab ära 39% andmete varieeruvusest.



Joonis 23: Päevase sammuga ruumientroopia prognoosjäägi standardhälbe seos GNSS andmetest arvatud ruumientroopiaga

Vaadeldes tegevusraadiust pole kasutatavad andmed normaaljaotusega, mis võib tekitada regressioonanalüüsi puhul probleeme, sest alg andmete normaaljaotus on üheks eelduseks. Tegevusraadiuse puhul mõjutavad sarnased vead täpsust nagu liikumisdistantsi puhul. Sarnaselt liikumisdistantsile ei ole mudel kuni 1km täpne, sest vigade osakaal mõjutab liigselt tulemust.

Kuna ruumientroopia regressioonimudelid on kasutatud parameetritest kõige kehvema andmete variatsiooni kirjeldamisega, siis võrdleme leitud regressioonimudelite standardiseeritud prognoosjääke, et vaadelda, millised parameetrid mõjutava üle- ja alahindamist. Kõige tugevam Spearmani korrelatsiooni seos on GNSS ruumientroopiaga. Mida suurem on vaadeldava päeva entroopia, seda rohkem alahindame

mobiilpositsioneerimise andmetega tegelikku ruumientroopiat. Antud seos tuleb välja vaid päevase sammuga. Nädalase sammuga ei ole olulisus piisav, et seda oleks võimalik kasutada. Mobiilpositsioneerimise punktide arvu seose tugevus on nii päevase kui ka nädalase sammuga on keskmine. Seos on positiivne ning olulisus $p > 0,001$. Sellest saame järeldada, et mida rohkem on mobiilpositsioneerimise punkte, seda suurem on tõenäosus ruumientroopia ülehindamiseks. Suurem punktide arv viitab tihedamale mastide vahetusele, mis viib omakorda entroopia ülehindamiseni. Samuti on mobiilpositsioneerimise ruumientroopia keskmise seosetugevusega korreleeruv ruumientroopia prognoosnäitude standardhälbega. Seosest saab välja lugeda, et mida suurem on mobiilpositsioneerimiselt arvatud ruumientroopia, seda suurema tõenäosusega on see ülehinnatud. Seos on täpselt vastupidine GNSS andmetelt arvatud ruumientroopiaga.

Viimasena vaatame ajalise entroopia ja ruumientroopia prognoosjääkide vahelist korrelatsiooni. Seos kahe parameetri vahel on oluline vaid päevase sammu korral ning seos on väga nõrk. Prognoosimise viga antud regressioonimudeli põhjal on liialt suur, et seda oleks võimalik kasutada ainukese kvaliteedi hindamise tunnusena. Samuti võib kvaliteedi hindamisel kasutada ajalist kattuvust, mille seose tugevus oli 0,43 ning mis võimaldab hinnata sarnaselt ajalisele entroopiale ruumientroopia tulemusi. Kuigi ajalise entroopia seos mobiilsusparameetrite suhtes oli nõrk, tuleks ajalist entroopiat uurida kõnetoimingute andmetega. Sellisel juhul oleks ajalise entroopia varieeruvus suurem ning annaks parema ülevaate, kuidas parameeter toimib mittetäielike andmete korral.

Töö hüpotees, ei leidnud kinnitust. Tulemustest selgus, et ajaline entroopia suurendab tõenäosust, et mobiilsusparameetrid: ruumientroopia ja liikumisdistsants hinnatakse üle. Autori arvates võib kõige olulisem põhjus olla ajalise entroopia tugev seos mobiilpositsioneerimise punktide arvuga, mis omakorda põhjustab tulemuste ala ning ülehindamist. Antud seosed said töös kinnituse. Teiseks probleemiks võib-olla andmete vähene varieeruvus, mis osaliselt on tingitud valimi suurusest ja mobiilpositsioneerimis andmetest, mis on täielikud.

Antud töös on kasutatud väikest valimit ning sellest tingituna kajastavad andmed vaid tiheasustusalasid. Yousense andmete kogumise meetodit ja kvaliteeti pole kasutajal võimalik muuta, seega kehtivad antud seosed kõigi YouSense rakenduse poole kogutud andmetele. Järgmised tööd võiksid vaadelda mobiilivõrgu tiheduse mõju mobiilsusparameetritele. Töös kasutatud kogu mobiiliside andmestiku asemel võiks vaadelda eraldi kõnetoiminguid ja mobiilset andmesidet ning leida, kas antud töös leitud seosed toimivad ka väiksemate ning

ruumiliselt kallutatud mobiilside andmete korral. Mobiilsusparameetritest võiks täpsemalt uurida olulisemate paikade tuvastamise täpsust.

5. Kokkuvõte

Magistritöö eesmärgiks oli koostada metoodika võrdlemaks GNSS andmeid ning mobiilpositsioneerimise andmeid kasutades selleks mobiilsusparameetreid. Teiseks uurida, kuidas mobiilpositsioneerimise andmete ajaline jaotus mõjutab ruumiparameetreid. Töö ainsaks hüpoteesiks oli: mobiilpositsioneerimise andmetest saadud indiviidi mobiilsusparameetrid on seda vähem kallutatud, mida ühtlasemalt on mobiilpositsioneerimis andmed jaotunud. Ajalist jaotumise hindamiseks kasutati ajalist entroopiat. Käesolevas töös on kasutatud YouSense rakenduse poolt kogutud mobiilpositsioneerimise andmeid ning GNSS andmeid. Mõlemad andmed tuli puhastada andmevigadest filtreerides mittevajalikud andmed välja. GNSS andmed grupeeriti peatuspaiga alusel. Saadud peatuspaigad grupeeriti omakorda DBSCAN ehk tihedusel põhineva algoritmide abil kokku. Antud analüüsi käigus arvutati kolm enam kasutusel olevat mobiilsusparameetrit: liikumisdistsants, ruumientroopia ja tegevusraadius. Peale nende parameetrite arvutati andmete ajaline kattuvus ja kirjete hulk. Parameetrid arvutati päevase ja nädalase sammuga nii YouSense GNSS andmetele kui ka mobiilpositsioneerimise andmetele. Lisaks arvutati mobiilpositsioneerimise andmete põhjal ajaline entroopia, mis näitab kui ühtlaselt on andmed ajaliselt jaotunud vaadeldava aja perioodi lõikes. Saadud tulemusi analüüsiti Spearmani korrelatsiooni ja regressioonanalüüsiga.

Käesolevas töös leiti tugev seos mobiilpositsioneerimise andmete ja GNSS andmete vahel nii liikumisdistsantsi kui ka tegevusraadiuse vahel ning keskmise tugevusega seos ruumientroopia arvutatuna kahest erinevast andmestikust. Tulemused on vastavuses varasemate uuringutega. Ajaline entroopia korreleerus positiivselt väga nõrgalt vaid ruumientroopia ja liikumisdistsantsi jäägi standardhällbega. Samuti leiti seosed punktide arvu ja liikumisparameetri prognoosjääkide vahel. Mobiilpositsioneerimise punktide arvu kui ka ruumientroopia abil on võimalik hinnata, kas tulemus on üle- või alahinnatud.

Töö hüpotees, et ajaline entroopia suurenedes mobiilsusparameetrite vead vähenevad, ei leidnud kinnitust. Nii andmete arvulise kui ajalise mahu kasvades suureneb tehtav viga. Vaatamata sellele, et ajalise entroopia seos mobiilsus parameetrite suhtes ei olnud tugev tuleks ajalist entroopiat uurida kasutades selleks suuremat valimit ja kõnetoimingute andmeid.

6. Impact of the time distribution of mobile positioning data on mobility parameters

Taavi Pipar

Summary

The aim of this master's thesis was to develop a methodology for comparing the data of the YouSense GNSS and mobile positioning data using the mobility parameters. The research hypothesis was that the individual mobility parameters from the mobile positioning data are with fewer errors if the time distribution data of the mobile positioning was more evenly distributed. YourSense mobile positioning and GNSS data were used in this research. The data were cleared. Stops were found from GNSS data. detected stop points were grouped by using the DBSCAN algorithm. Mobility parameters: distances, movement entropy and radius of gyration were calculated from both data sets. In addition, time coverage and the number of records were calculated. All parameters were calculated daily and weekly interval. In addition, time entropy was calculated, based on mobile positioning data, which shows how consistently the data is distributed over time. The results were analysed by Spearman correlation and regression analysis. In the present work, a very strong correlation was found between the mobile positioning and GNSS data, using distance and radius of gyration. The correlation between the movement entropy in two data sets was an average strength. Further, the correlation between movement entropy error and temporal entropy, time coverage, and a number of records was analysed. The study revealed that temporal entropy correlates positively but weakly with movement entropy errors. It also correlated positively with a number of records in mobile positioning data and time coverage with average strength. Based on these results, it is possible to evaluate whether the result is overestimated or overestimated. The study hypothesis that the time entropy increases while mobility errors decline was not confirmed. feature studies should look into a bigger dataset and use temporal entropy with passive mobile positioning data.

7. Tänuavaldused

Soovin tänada juhendajat Anto Aasat mitmekülgse abi ja nõuannete ning pühendatud aja eest. Tänan ka igati toeks olnud peret ja lähedasi.

Kasutatud kirjandus

- Blondel, V.D., Decuyper, A., Krings, G., 2015. A survey of results on mobile phone datasets analysis. *EPJ Data Sci.* 4, 10. <https://doi.org/10.1140/epjds/s13688-015-0046-0>
- Calabrese, F., Di Lorenzo, G., Liu, L., Ratti, C., 2011. Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area. *IEEE Pervasive Comput.* 10, 36–44.
- Chen, G., Hoteit, S., Viana, A.C., Fiore, M., Sarraute, C., 2017a. Enriching Sparse Mobility Information in Call Detail Records (report). INRIA Saclay - Ile-de-France.
- Chen, G., Viana, A.C., Sarraute, C., 2017b. Towards an adaptive completion of sparse Call Detail Records for mobility analysis, in: 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). Presented at the 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pp. 302–305. <https://doi.org/10.1109/PERCOMW.2017.7917577>
- Cich, G., Knapen, L., Bellemans, T., Janssens, D., Wets, G., 2015. TRIP/STOP Detection in GPS Traces to Feed Prompted Recall Survey. *Procedia Comput. Sci.*, The 6th International Conference on Ambient Systems, Networks and Technologies (ANT-2015), the 5th International Conference on Sustainable Energy Information Technology (SEIT-2015) 52, 262–269. <https://doi.org/10.1016/j.procs.2015.05.074>
- Frias-Martinez, V., Virseda, J., 2012. On the Relationship Between Socio-economic Factors and Cell Phone Usage, in: Proceedings of the Fifth International Conference on Information and Communication Technologies and Development, ICTD '12. ACM, New York, NY, USA, pp. 76–84. <https://doi.org/10.1145/2160673.2160684>
- González, M.C., Hidalgo, C.A., Barabási, A.-L., 2008. Understanding individual human mobility patterns. *Nature* 453, 779–782. <https://doi.org/10.1038/nature06958>
- Hoteit, S., Secci, S., Sobolevsky, S., Ratti, C., Pujolle, G., 2014. Estimating human trajectories and hotspots through mobile phone data. *Comput. Netw.* 64, 296–307. <https://doi.org/10.1016/j.comnet.2014.02.011>
- Iovan, C., Olteanu-Raimond, A.-M., Couronné, T., Smoreda, Z., 2013. Moving and Calling: Mobile Phone Data Quality Measurements and Spatiotemporal Uncertainty in Human Mobility Studies, in: Geographic Information Science at the Heart of Europe, Lecture Notes in Geoinformation and Cartography. Springer, Cham, pp. 247–265. https://doi.org/10.1007/978-3-319-00615-4_14
- Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., Varshavsky, A., 2011. Identifying Important Places in People's Lives from Cellular Network Data, in: Pervasive Computing, Lecture Notes in Computer Science. Presented at the International Conference on Pervasive Computing, Springer, Berlin, Heidelberg, pp. 133–151. https://doi.org/10.1007/978-3-642-21726-5_9
- Karsai, M., Kaski, K., Barabási, A.-L., Kertész, J., 2012. Universal features of correlated bursty behaviour. *Sci. Rep.* 2, 397. <https://doi.org/10.1038/srep00397>
- Laurila, J.K., Gatica-Perez, D., Aad, I., Blom, J., Bornet, O., Do, T.M.T., Dousse, O., Eberle, J., Miettinen, M., 2013. From big smartphone data to worldwide research: The Mobile Data Challenge. *Pervasive Mob. Comput.*, Mobile Data Challenge 9, 752–771. <https://doi.org/10.1016/j.pmcj.2013.07.014>
- Linnap, M., Rice, A., 2014. Managed Participatory Sensing with YouSense. *J. Urban Technol.* 21, 9–26. <https://doi.org/10.1080/10630732.2014.888216>

- Lu, S., Fang, Z., Zhang, X., Shaw, S.-L., Yin, L., Zhao, Z., Yang, X., 2017. Understanding the Representativeness of Mobile Phone Location Data in Characterizing Human Mobility Indicators. *ISPRS Int. J. Geo-Inf.* 6, 7. <https://doi.org/10.3390/ijgi6010007>
- Naboulsi, D., Fiore, M., Ribot, S., Stanica, R., 2016. Large-Scale Mobile Traffic Analysis: A Survey. *IEEE Commun. Surv. Tutor.* 18, 124–161. <https://doi.org/10.1109/COMST.2015.2491361>
- Pestre, G., Letouzé, E., Zagheni, E., 2016. The ABCDE of Big Data. Assessing Biases in Call-Detail Records for Development Estimates. *Annu. Bank Conf. Dev. Econ.*, June.
- Ranjan, G., Zang, H., Zhang, Z.-L., Bolot, J., 2012. Are Call Detail Records Biased for Sampling Human Mobility? *SIGMOBILE Mob Comput Commun Rev* 16, 33–44. <https://doi.org/10.1145/2412096.2412101>
- Schulz, D., Bothe, S., Körner, C., 2012. Human mobility from gsm data-a valid alternative to gps, in: *Mobile Data Challenge 2012 Workshop*, June. pp. 18–19.
- Song, C., Qu, Z., Blumm, N., Barabási, A.-L., 2010. Limits of Predictability in Human Mobility. *Science* 327, 1018–1021. <https://doi.org/10.1126/science.1177170>
- Sun, Q., Ahas, R., Aasa, A., Wan, W., Yuan, C., 2017. GPS data cleaning and analysis based on YouSense mobile application.
- Trestian, I., Ranjan, S., Kuzmanovic, A., Nucci, A., 2009. Measuring Serendipity: Connecting People, Locations and Interests in a Mobile 3G Network, in: *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, IMC '09*. ACM, New York, NY, USA, pp. 267–279. <https://doi.org/10.1145/1644893.1644926>
- Vent, K., 2014. Inimese tegevuskohtade leidmine nutitelefonipõhiste käitumisandmestike alusel (Thesis). Tartu Ülikool.
- Wesolowski, A., Eagle, N., Noor, A.M., Snow, R.W., Buckee, C.O., 2013. The impact of biases in mobile phone ownership on estimates of human mobility. *J. R. Soc. Interface* 10, 20120986. <https://doi.org/10.1098/rsif.2012.0986>
- Yuan, Y., Raubal, M., Liu, Y., 2012. Correlating mobile phone usage and travel behavior – A case study of Harbin, China. *Comput. Environ. Urban Syst., Special Issue: Geoinformatics* 2010 36, 118–130. <https://doi.org/10.1016/j.compenvurbsys.2011.07.003>
- Zhang, D., Huang, J., Li, Y., Zhang, F., Xu, C., He, T., 2014. Exploring Human Mobility with Multi-source Data at Extremely Large Metropolitan Scales, in: *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking, MobiCom '14*. ACM, New York, NY, USA, pp. 201–212. <https://doi.org/10.1145/2639108.2639116>
- Zhao, Z., Shaw, S.-L., Xu, Y., Lu, F., Chen, J., Yin, L., 2016. Understanding the bias of call detail records in human mobility research. *Int. J. Geogr. Inf. Sci.* 30, 1738–1762. <https://doi.org/10.1080/13658816.2015.1137298>

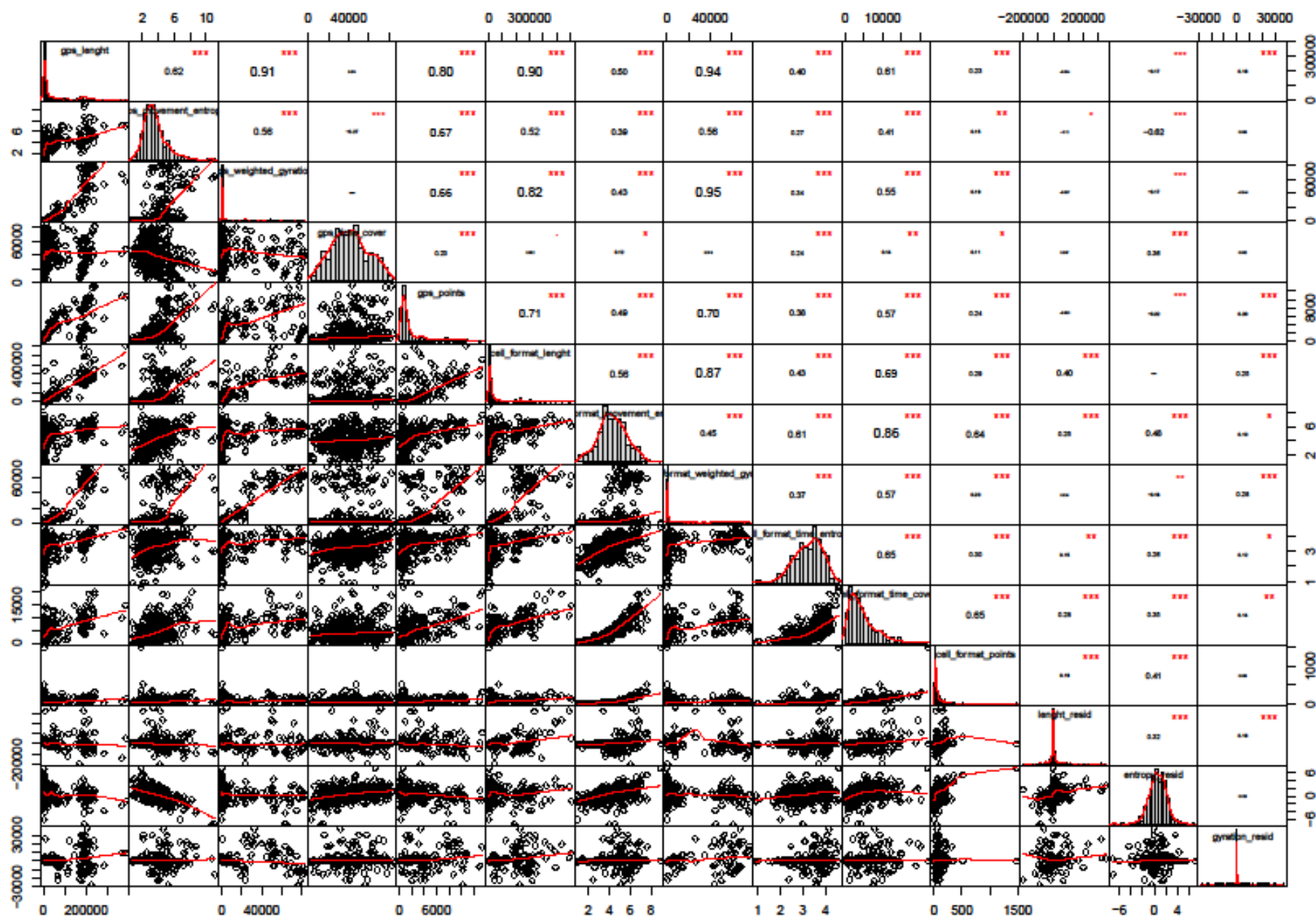
Internetiallikad

OpenCellID, n.d. URL <https://www.opencellid.org/#zoom=16&lat=37.77889&lon=-122.41942> [15.5.18].

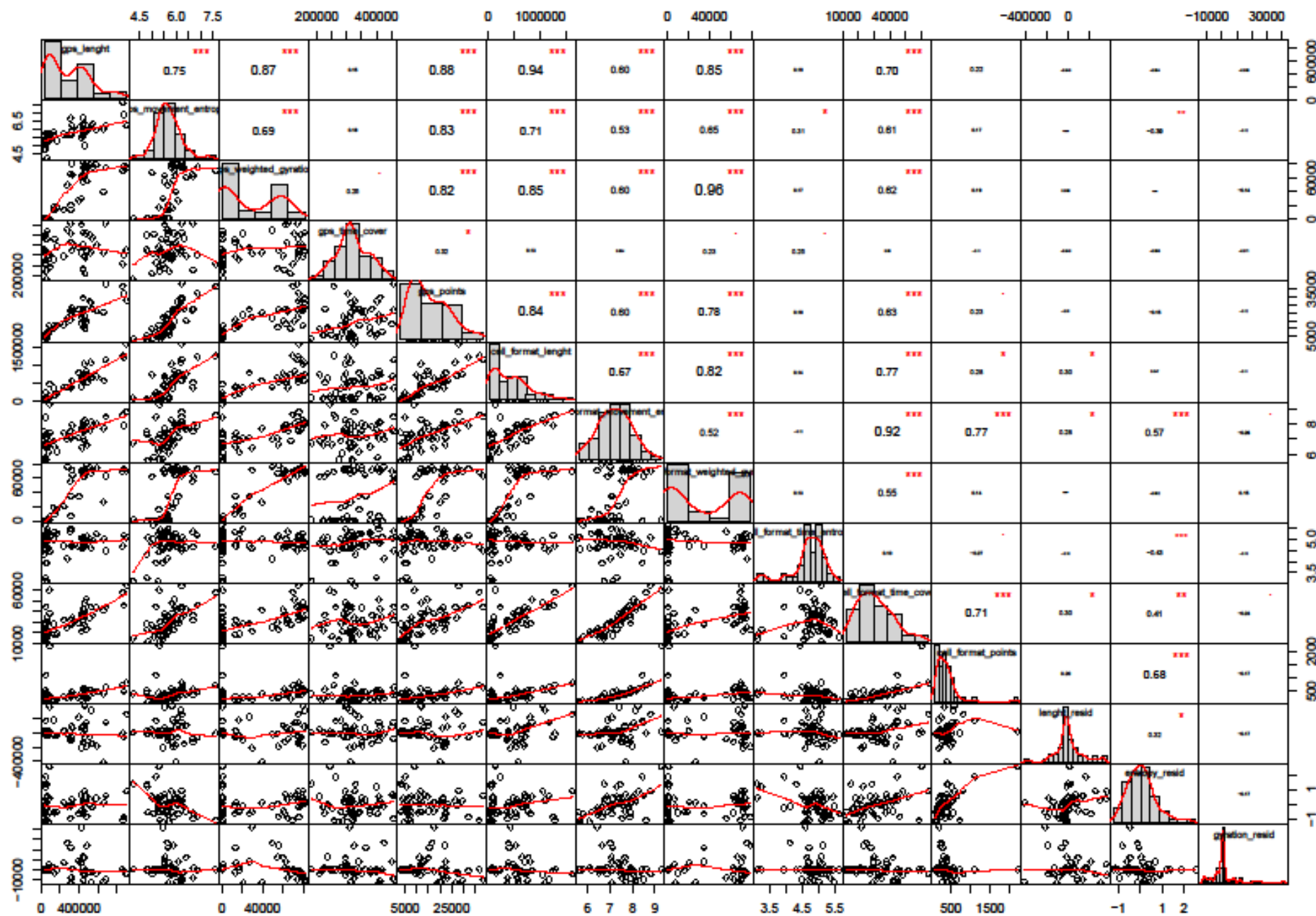
YouSense, n.d. URL <http://positioner.ut.ee/dashboard/info/> [11.5.18].

Telia leviala, n.d. URL <https://www.telia.ee/et/era/mobiil/leviala> [11.5.18]..

Lisa 1



Lisa 2



Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Taavi Pipar,

annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

„Mobiilpositsioneerimise andmete ajalise jaotuse mõju mobiilsuse parameetritele“, mille juhendaja on Anto Aasa,

reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni; üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 28.05.2018